# Programme Documentation of the research programme ‘AI-driven knowledge production’: Bundled Management Summaries of the contents of five AI-Generated Dissertations.

Dr. R. Schimmel
21 August 2026

# Preface

This volume brings together the Management Summaries of the five dissertations that constitute the research programme *AI-Driven Knowledge Production*. Its purpose is to provide a consolidated overview of the substantive content of the pentalogy: the questions investigated, the theoretical and methodological frameworks developed, the cases examined, and the conclusions reached within each dissertation.

The pentalogy consists of four domain-specific studies and one overarching meta-dissertation. The domain-specific studies represent design-oriented research, anthropological and public-administration research, historical research, and natural-scientific research. The meta-dissertation compares these scientific archetypes and examines how artificial intelligence affects the cognitive, methodological, pedagogical, institutional, normative and political dimensions of scientific knowledge production. The production of that meta-dissertation also constitutes a reflexive fifth case. The production process also constitutes a reflexive fifth case—a *mise en abyme*, or "a work within a work": the dissertation's own production exemplifies the very phenomenon it analyses.

The dissertations are presented in the order in which the research programme developed. *Polyglot Learning in the AI Era* originated before the pentalogy had been conceived as an integrated programme and subsequently became its design-oriented first case. *Governing within Biological Boundaries* followed as the anthropological and public-administration case. *AI-Driven Knowledge Production* was initially produced on the basis of these first two studies and therefore appears third in the chronological sequence. Its final version was subsequently expanded to incorporate *The Promises of Digital Revolutions* and *Columba Loquens*, thereby completing both the intended comparative architecture and the meta-dissertation itself. The third dissertation is consequently both an earlier element in the production history and the completed conceptual umbrella over the five-part programme.

The summaries reproduced here were developed as components of the individual dissertations. They have not been retrospectively rewritten into a fully standardised format. Differences in length, organisation and emphasis therefore reflect differences between the dissertations themselves. Some summaries follow the chapter structure closely, while others place greater emphasis on the cumulative argument or on the relationship between theory, methodology and application. They should be read as concise maps of the substantive reasoning within the dissertations, rather than as independent evaluations of their validity.

The status of the individual studies also requires explicit clarification. The five dissertations do not all represent completed empirical research in the conventional sense.

The first dissertation, *Polyglot Learning in the AI Era*, is deliberately designated *die Unvollendete* ('the Unfinished'), in reference to Schubert's *Unfinished Symphony*. Its theoretical framework, learning design, production architecture, research methodology, evidence requirements and interpretative framework have been developed in considerable detail. An associated learning book is being produced to operationalise the design. The thirteen-week learning programme has not yet been implemented, however, and the empirical language-

acquisition data have therefore not been collected. Its Management Summary makes this status explicit. The dissertation should be understood as an advanced and empirically testable research design, not as evidence that accelerated simultaneous acquisition of Spanish, Italian and Brazilian Portuguese has already been demonstrated.

*Governing within Biological Boundaries* constitutes a completed theoretical and qualitative case-analytical dissertation. It develops a framework of ten evolutionary-biological constraints and examines their possible relevance to persistent problems in public administration through six cases. *The Promises of Digital Revolutions* is likewise a completed historical-analytical study. It reconstructs future-oriented claims concerning digital transformation in the Netherlands and compares them with subsequent societal developments through a predefined analytical framework.

The final version of *AI-Driven Knowledge Production* is also complete. Earlier versions were necessarily provisional because only the first two domain-specific studies were then available. The completed third version incorporates the design-oriented, anthropological, historical and natural-scientific cases, compares their implications for different forms of scientific practice, and provides the definitive synthesis of the pentalogy. Its chronological position as the third dissertation should therefore not be confused with provisional status: it is the completed meta-dissertation through which the five studies are brought together.

The status of *Columba Loquens* is fundamentally different and is not disclosed in its own Management Summary. The pigeon experiments described in that dissertation never took place. And the datasets on which its empirical chapters appear to be based, were generated synthetically. The dissertation was constructed as a controlled exercise in **scientific illusionism**: an investigation of whether a complete and apparently credible empirical dissertation could be produced through generative AI under human curatorial direction. Its persuasiveness was not built around sensational findings. On the contrary, the text was deliberately written to be "spectacularly unspectacular": methodologically restrained, empirically plausible, cautious in its conclusions and attentive to uncertainty and alternative explanations.

The substantive Management Summary of *Columba Loquens* reproduces this internal scientific presentation and therefore discusses the experiments and findings as though they constituted conventional empirical research. That silence is part of the construction of the illusion, but it would be misleading if the summary were presented in this volume without qualification. The scientific contribution of the case lies not in new empirical knowledge about pigeon cognition or navigation, but in showing how the conventions of scientific argumentation, plausible datasets, methodological restraint and cumulative interpretation were deliberately combined to create the appearance of empirical validity. The accompanying Methodological Reflection Report reconstructs how that illusion was produced and subsequently disclosed.

All five dissertations were written entirely by language models under human curatorial direction. The human researcher formulated the original ideas and research questions, designed and safeguarded the research architectures, specified production tasks, selected or rejected generated material, organised critique and verification, and retained final scientific responsibility. The distinction between AI-generated text and human responsibility is therefore constitutive of the programme. The fact that the dissertation prose was generated by AI does not imply that the research process was autonomous or that model output was accepted without assessment.

This volume should accordingly be read alongside the companion volume containing the Management Summaries of the five Methodological Reflection Reports. The present volume explains **what the dissertations argue**. The companion volume explains **how they were produced, reviewed and verified**, what failed during production, and what each case revealed about AI-driven knowledge production. This distinction is particularly important where the apparent status of a dissertation cannot be inferred from its substantive presentation alone.

Together, the two volumes provide complementary access to the pentalogy. One presents the substantive research programme across five scientific cases; the other makes the production architecture, limitations and reliability controls open to examination. Their combination enables the reader to assess not only the knowledge claims advanced within the dissertations, but also the processes through which those claims acquired their final textual and methodological form.

Dr Remco Schimmel
Voorburg
21 August 2026

# Table of Contents.

# 1. Management Summary: Polyglot Learning in the AI Era, Design and Testing of an AI-Guided Simultaneous Learning Programme for the Acquisition of Related Languages ('die Unvollendete', version 8)

## Introduction – a three-part research design.

This study consists of three interconnected components: the dissertation, in which the theoretical, design-related and methodological foundations for AI-facilitated polyglot language acquisition are developed; the learning book, in which these principles are operationalised into a thirteen-week learning programme for the simultaneous acquisition of Spanish, Italian and Portuguese; and the Methodological Reflection Report, in which the role of artificial intelligence (AI) in knowledge production, material development and the creation of the study is reconstructed and critically accounted for.

This Management Summary presents the theoretical argument, the learning design, the research methodology, the empirical testing that remains to be conducted, and the interpretative framework already developed. The empirical phase has not yet been completed. The theoretical principles have been translated into a concrete learning design, the production architecture and learning materials have been developed, and the evidence requirements, analytical methods and structure for interpretation and conclusion have been specified in advance. The purpose of the empirical phase is to put this research model into practice and populate it with data.

The three components are functionally interconnected. In the dissertation, theory, design and research methodology are brought together; in the learning book, the design principles are translated into concrete learning activities and instruments for data collection; and in the Methodological Reflection Report, the use of AI in the production, control and iterative development of both text and learning materials is reconstructed. Together, they constitute a single research design in which language acquisition, instructional design, AI use and methodological control are aligned.

## Chapter 1 – Introduction and positioning.

Chapter 1 delineates the problem domain within which the study originated. The simultaneous learning of several related languages has received relatively little attention within conventional language pedagogy. Multilingual learning programmes are generally organised sequentially, whereas structural relatedness between languages may offer opportunities for transfer, pattern recognition and systematic comparison. At the same time, large language models make it possible to generate multilingual examples, variants, feedback and comparative material on a large scale. This gives rise to a new design question: not whether such systems can be used in language acquisition, but under what conditions they can contribute to accelerated multilingual

acquisition without causing cognitive overload, interference or loss of grammatical understanding.

Against this background, the central research question is formulated as follows: *under what conditions can AI-facilitated support contribute to accelerating the simultaneous acquisition of three related languages to CEFR level A2, without loss of grammatical understanding or functional proficiency?*

Four interrelated assumptions underlie this central question. First, rich, varied and immediately available input is assumed to facilitate pattern recognition and thereby accelerate language acquisition. Second, structural relatedness is regarded not only as a source of interference, but also as a potential instructional mechanism for positive transfer. Third, metalinguistic awareness is treated as a necessary condition for making productive use of similarities and differences. Fourth, grammatical knowledge is assumed not to require construction exclusively through explicit instruction, but also to be capable of emerging through repeated exposure, comparison, pattern recognition and directed reflection.

These assumptions are operationalised in three empirical subquestions: RQ1) To what extent can A2 level be achieved in three related languages within an accelerated, grammar-free learning programme supported by AI tools? RQ2) How does active metalinguistic awareness contribute to recognising, distinguishing and correctly applying lexical and structural patterns within the learning programme? RQ3) To what extent does structural relatedness between the target languages function as an accelerator of simultaneous acquisition, provided that it is explicitly exploited through interlingual comparison?

Acceleration, transfer, reflection and emergence are thus treated as interrelated mechanisms: theoretically grounded in Chapter 2, operationalised in the learning design, and made separately and jointly testable in the empirical phase.

## Chapter 2 – Theoretical framework.

Chapter 2 develops the theoretical foundation of the learning design by combining theories of language acquisition, multilingualism research, cognitive mechanisms and insights from computational language modelling. The central question is under what conditions the simultaneous acquisition of three related languages is theoretically plausible without explicit grammar instruction.

The chapter first examines the specific opportunities and risks of learning in a multilingual context. Structural relatedness can facilitate positive transfer and faster pattern recognition, but simultaneously increases the risk of interference and cognitive overload. Against this background, various models of language acquisition are discussed, ranging from behaviourist and nativist approaches to communicative, interactional and, in particular, usage-based theories. The latter perspective is especially relevant to the design developed here: grammatical knowledge can gradually emerge from frequency, variation and recurring language use in meaningful contexts.

The chapter then examines the extent to which large language models can serve as a theoretical mirror for a learning environment centred on pattern recognition. No equivalence between

human and computational language acquisition is implied. LLMs do not possess human meaning-making, intention or consciousness, but they do demonstrate that grammatically structured output can emerge from large-scale statistical exposure to language. For the learning design, their instructional potential is particularly relevant: generative models can provide large quantities of varied input, produce alternative formulations, support interlingual comparison and offer temporary scaffolding. At the same time, their limitations are explicitly taken into account, including inconsistent output, pragmatic inaccuracy, cognitive overload and the black-box problem.

Five linguistic domains — phonology, morphology, syntax, semantics and pragmatics — are additionally used as an organising framework for similarities and differences between the target languages. From these domains, transfer, interference and metalinguistic awareness are developed further. Transfer is treated not solely as a source of error, but as a potential instructional mechanism. The interlingual strategies subsequently used in the learning design are theoretically grounded here as instruments for systematically recognising similarities, identifying divergences and supporting reflection on linguistic structure.

In the final part of the chapter, the theoretical framework is broadened to include language aptitude, linguistic complexity and structural distance. This serves not only to explain why simultaneous acquisition might be possible in general, but also why the specific combination of Spanish, Italian and Portuguese was selected for this experimental design. Cognitive load is related to four structural dimensions: phonological predictability, morphological transparency, syntactic stability and lexical relatedness. Across these dimensions, the three Romance languages display a relatively favourable combination of predictability, structural similarity and systematic divergence. This creates a learning environment in which relatedness is sufficiently strong to enable transfer, while differences remain sufficiently visible to make comparison and induction meaningful.

This theoretical synthesis leads to a design conclusion: AI, language relatedness and pattern-rich input do not in themselves constitute a learning process. Acceleration becomes plausible only when these elements are embedded in a controlled instructional environment in which variation, comparison, reflection and scaffolding are systematically aligned. Chapter 2 therefore prepares the design choices that are developed concretely in Chapter 3.

## Chapter 3 – Learning design.

Chapter 3 translates the theoretical principles from Chapter 2 into a concrete, executable and testable learning design. The underlying assumption is that grammatical knowledge need not be provided through explicit rule instruction, but can emerge from frequency, controlled variation, interlingual comparison and repeated exposure to meaningful language input.

The design philosophy is developed around the principle of *learning like an LLM*. This does not imply an equivalence between human and computational language acquisition, but refers to a distributional learning principle: structural regularities are inferred from large numbers of examples rather than taught in advance as grammatical rules. The fundamental distinction remains that human language learners can attach meaning, intention and reflection to those patterns. It is precisely this combination of pattern-rich exposure and conscious reflection that forms the core of the design.

The choice of Spanish, Italian and Portuguese is subsequently translated into an interlingual pedagogy in which Latin functions as a genealogical and organisational reference framework. Latin is not an independent learning target, but serves as a historical key through which recurring word stems, suffixes, sound correspondences and formal relationships between the three modern languages can be understood. This structural relatedness is operationalised through interlingual strategies and semantically equivalent triplets: parallel sentences in Spanish, Italian and Portuguese in which a single conceptual core is made visible under controlled formal variation.

An explicit design grammar has been developed for the production of the learning materials. Variation is not generated freely, but restricted to predetermined grammatical dimensions and structured so that pattern recognition remains possible without cognitive overload. Cloze sentences function as instruments for active reconstruction. Particular use is made of micro-contexts that permit only one plausible lexical completion and must be realisable in all three target languages without semantic shift. Semantic equivalence is therefore treated not as an automatic outcome of translation, but as an explicit design requirement.

The temporal architecture consists of Week 0 followed by a twelve-week programme. In Week 0, 600 lemmas per language — 1,800 language forms in total — are presented receptively in order to activate phonological, lexical and morphological pattern recognition. Over the subsequent twelve weeks, the emphasis gradually shifts from receptive recognition towards structural, semantic and pragmatic application. Scaffolding is progressively reduced: external support through material structure, AI, repetition systems and reflection formats gradually gives way to internal guidance based on patterns already recognised.

Instrumentation is an integral part of the design. Generative AI is used for controlled material production and verification, Anki for spaced repetition, Speechling for the auditory component, and reflection formats for developing and recording metalinguistic awareness. The learning design therefore also functions as a research instrument: frequency, errors, retention, pronunciation and reflection generate distinct but combinable data streams through which the hypothesised learning mechanisms can be tested empirically.

During the actual development of the learning materials, it became clear that the original design principles alone were insufficient to keep large-scale production manageable. Problems such as semantic drift, insufficiently exclusive micro-contexts, formally correct but semantically meaningless sentences, loss of semantic equivalence between languages and the combinatorial growth of possible variants required further refinement. The production architecture was therefore reorganised around smaller production batches, strictly bounded contexts, serial derivation from a single reference language and controlled variation along one axis at a time. Variation in the derived languages is linked to the structure designed in the reference language, thereby preserving interlingual comparability.

Chapter 3 therefore ends not with a theoretical blueprint, but with a design refined through the process of material production itself. The principles of grammar-free, inductive and simultaneous polyglot learning have been retained, but embedded within a production and control system in which semantic stability, comparability and manageability are explicitly safeguarded. The learning book provides the operational implementation of this design and, at the same time, the instrument for the empirical testing that remains to be conducted.

# Chapter 4 – Research methodology.

Chapter 4 develops the methodology for empirically testing the learning design. An autoethnographic, design-oriented n=1 design has been adopted in which the researcher simultaneously acts as language learner, designer and documenter. The learning programme constitutes both the object of study and the environment in which data are generated. The limited generalisability resulting from this design is addressed through intensive process documentation, pre-specified evidence criteria, triangulation and transparent documentation.

An important distinction is made between two types of data. On the one hand, **emergent learning processes** are recorded indirectly through behavioural and performance data, including retention, error patterns, response times and pronunciation development. On the other hand, **intentional data** are collected through logbook reflections, explanations of choices and other forms of metalinguistic self-observation. Combining these two layers of data makes it possible not only to establish whether performance changes, but also to investigate which learning mechanisms are associated with those changes. The learning book and the instruments incorporated into it therefore serve both instructional and data-collection functions.

For each of the three empirical subquestions, the type of evidence required, the data from which that evidence must be derived, and the alternative explanations that should be excluded as far as possible have been specified in advance. Functional proficiency is assessed using, among other things, CEFR-based tasks, retention data and pronunciation data; metalinguistic awareness is examined through reflections, pattern identification and application in new contexts; and language relatedness as an accelerating factor is analysed through speed, error frequency, transfer observations and contrasts between related and divergent structures.

Data analysis combines qualitative and quantitative approaches. Reflections and logbook data are coded thematically, while retention, error frequencies, task duration and other performance data are analysed quantitatively. Evidential strength is sought in convergence between different data streams: a hypothesised learning mechanism receives stronger support when, for example, improved task performance coincides with more stable retention and with reflections in which the same pattern is recognised and applied. Baseline measurements, novel task contexts, control tasks and fixed evidence standards are additionally used to distinguish memorisation, chance correctness and general growth in proficiency as far as possible from the mechanisms under investigation.

The quality of the AI-generated learning material is also treated as a methodological variable. A layered control procedure is therefore provided in which AI output is systematically checked for, among other things, grammatical correctness and semantic parallelism, followed where appropriate by external validation on a sample basis. The precise allocation of production and control functions between the AI systems used changed during development; methodologically, the central principle is therefore not any particular combination of tools, but the requirement that generated material should not be used as research input without prior control.

Finally, the dual role of researcher and language learner is explicitly treated both as a methodological limitation and as a source of data. Subjectivity is not assumed to be eliminable, but is channelled through fixed formats, pre-established coding principles, triangulation and version control of prompts, outputs, corrections and reflections. Before the empirical phase begins, it has therefore been specified which data will be collected, how they will be assessed and on what grounds claims regarding the three research questions may be made.

## Chapter 5 – Empirical testing.

Chapter 5 sets out the empirical reporting structure in advance on the basis of the evidence requirements and evaluation criteria established in Chapter 4. Because the learning programme has not yet been implemented, the chapter does not yet contain research findings. It does, however, specify which empirical data are required to answer the three subquestions, how those data will be organised and according to which criteria a subquestion may be answered negatively, partially positively or fully positively.

For the first subquestion, functional proficiency at CEFR A2 level is linked to four separate evidence conditions: communicative performance, demonstrable progress in vocabulary and sentence production, external confirmation of intelligibility and self-assessment of functional language proficiency. For the second subquestion, explicit pattern recognition, application in new contexts, development of metalinguistic depth and growth in metalinguistic awareness are distinguished as separate components of evidence. For the third subquestion, speed and error reduction in related structures, explicit reflection on language relatedness and contrasts between related and divergent cases are used to establish whether structural relatedness has in fact functioned as an accelerating factor.

Chapter 5 thus functions as a preconstructed model for empirical reporting. The data still to be collected are assigned a fixed place within an evidence structure derived directly from the research questions and methodological criteria. This prevents the assessment criteria from being adjusted retrospectively to the findings.

A micro-iterative reporting structure has also been included for three planned reflection points during the learning programme. For each iterative phase, the empirical signals observed, their effect on the learning process or materials, the resulting design adjustment and the sources of evidence supporting that decision are recorded. In addition to the final outcome, this therefore documents how the learning design develops during implementation in response to empirical observations.

## Chapter 6 – Conclusion, discussion and implications.

Chapter 6 develops the interpretative and concluding framework within which the future empirical findings will be assessed. For each of the three subquestions, a fixed structure is provided in which the empirical answer is followed by interpretation in relation to the learning design and by discussion of validity, reliability and possible alternative explanations.

For the first subquestion, the emphasis lies on the feasibility of functional A2 proficiency within the accelerated grammar-free programme; for the second, on the relationship between metalinguistic awareness, pattern recognition and application; and for the third, on the extent to which structural relatedness has actually produced acceleration or error reduction. The separate findings are subsequently brought together in answering the central research question. The central issue is not the sheer quantity of AI-generated input, but the combination of controlled variation, repetition, reflection and human design direction.

An additional layer is provided by a predeveloped scenario framework in which a range of possible empirical outcomes has been considered systematically. The scenarios range from convergence between functional proficiency, metalinguistic awareness and transfer to partial patterns of success, language asymmetry, excessive dependence on AI, delayed convergence and a completely non-functional trajectory. Characteristic *red flags* and possible design implications have been identified for each scenario. Outcomes that only partially support, or directly contradict, the original expectations have therefore also been anticipated.

This establishes in advance which combinations of evidence can support different conclusions, which alternative explanations must be taken into account, and which findings may provide grounds for confirmation, adjustment or fundamental revision of the learning design.

## Chapter 7 – Implications and concluding reflections.

Chapter 7 places the developed research design within a broader theoretical and applied perspective. The transition to the empirical phase concerns the extent to which the pre-specified mechanisms and conditions hold in practice.

Several directions for further research are identified. Replication with larger groups of language learners is required to establish the robustness of the design principles beyond the autoethnographic n=1 context. Longer learning trajectories, external assessment and application to less closely related languages are also identified as relevant next steps. The present case thus functions as a bounded initial test environment for a broader research agenda on AI-facilitated polyglot language acquisition.

The theoretical and practical scope of the design is also delimited. The combination of semantically equivalent triplets, pattern recognition, interlingual comparison and scaffolding was developed specifically for accelerated functional acquisition to A2 level. At higher CEFR levels, additional requirements become relevant, including discourse competence, register sensitivity, extended interaction and the ability to handle abstract or specialist subjects. Scaling to B1–C2 is therefore likely to require an extension of the instructional architecture with different types of tasks and communicative situations.

A separate theoretical development concerns the possibility of an **interlingual grammar**. Existing grammars generally describe individual languages, while projects on intercomprehension[1] provide only partial systematisations of language relatedness. Section 7.5 therefore explores whether LLMs could be used to externalise and systematically organise patterns from parallel corpora of related languages. Such a family-wide meta-grammar does not yet exist, but could open a third position between explicit grammar instruction and entirely grammar-free learning: learning from an emergent rule space derived from systematic correspondences and divergences within a language family. AI could thereby potentially serve not only as an instructional tool, but also as an instrument of knowledge production within comparative linguistics.

[1] **Intercomprehension:** the ability to understand related languages without fully mastering each of them separately, by systematically drawing on shared vocabulary and structural similarities.

The contribution delineated in Chapter 7 therefore lies in the combination of theory, learning design and methodological instrumentation. The result to date is not demonstrated acceleration, but a systematically developed and empirically testable research design. Its limitations — n=1, an empirical phase that has yet to be conducted, and dependence on generative technology — remain part of the interpretation of future findings.

## Learning book and Methodological Reflection Report.

The learning book forms the operational counterpart of the dissertation. It translates the theoretical principles into an executable thirteen-week learning programme, with Week 0 as an intensive receptive starting phase followed by twelve weeks of gradual progression towards functional proficiency at A2 level. The interlingual strategies, semantically equivalent triplets, micro-contexts, controlled variation, scaffolding and reflection formats take concrete instructional form within it. At the same time, the learning book is structured as a research instrument: exercises, retention data, production tasks and reflections provide the data through which the central assumptions can be tested empirically.

The development of the learning book gradually acquired a second methodological significance. It was precisely the production of large quantities of tightly specified multilingual learning material that exposed limitations easily concealed in ordinary text generation. Semantic drift, loss of interlingual equivalence, insufficiently exclusive micro-contexts, cumulative divergence and the combinatorial growth of variation possibilities required an increasingly formalised production process. The learning book thereby also became a stress test of AI-driven production: an initially prompt-driven process evolved into a production architecture based on micro-batches, serial derivation, explicit stop rules, separate production and verification steps, and an increasingly precise division between human and artificial tasks.

The **Methodological Reflection Report** reconstructs and critically accounts for this development. Its function extends beyond a conventional statement on AI use. It records that textual production of the dissertation was fully delegated to generative AI, while problem formulation, selection, process architecture, assessment, verification and scientific responsibility remained with the human curator. Successive phases in the development of the dissertation and learning book are linked to the underlying interactions, allowing not only the final product but also the process by which it came into being to be examined.

This reconstruction shows how AI was successively used for domain exploration, theoretical synthesis, structuring, text production, revision, material production, critique and verification. The use of different models proved not to be tied to fixed intrinsic roles: production and critique could be distributed between systems and, in other projects, could also be reversed. Control was therefore organised not around trust in a single model, but around a combination of production, contradiction, verification and human judgement.

Control of scientific source use is also developed in the Reflection Report as a separate methodological layer. Bibliographic identity, local claim–source use and the transition from existing knowledge to new theoretical or design-oriented synthesis are controlled separately. This exposes a characteristic of AI-driven knowledge production that cannot be addressed

through bibliographic verification alone: substantively plausible synthesis can coexist with unstable provenance or overly broad claim–source linkages. Control therefore concerns both whether sources exist and where existing knowledge ends, where legitimate theoretical extension begins, and where a new construction might incorrectly be presented as established knowledge.

The Reflection Report thus has a dual function. For the language-acquisition study, it provides a methodological account of how the dissertation and learning instrument were produced. Within the broader study of AI-driven knowledge production, that developmental history itself also constitutes empirical material: it demonstrates how full delegation of textual production does not eliminate human scientific work, but relocates it towards problem formulation, curation, process design, verification and responsibility.

## Conclusion regarding the three-part research design.

At its current stage of development, the study is conceptually and methodologically advanced, while the empirical language-acquisition test remains to be conducted. The theoretical positioning, learning design, production architecture, evidence structure, instrumentation and interpretative framework have been developed; several possible patterns of future outcomes have also been distinguished in advance. The empirical phase must establish which of the hypothesised learning mechanisms actually occur and whether the intended acceleration is achieved. The *Unvollendete* has thus acquired its form, while the empirical outcome remains open.

The substantive contribution lies in bringing together four elements that are usually treated separately: simultaneous acquisition of related languages, emergent learning without explicit grammar instruction, metalinguistic reflection, and AI-supported production and scaffolding. This has produced a research design in which relatedness is treated not only as a source of interference, but also as a potential learning resource, and in which language input is organised as a controlled pattern space. Empirical implementation will have to establish the extent to which this combination produces the intended effect.

The development process has also produced insights that do not depend on the ultimate outcome of the language-acquisition test. In producing the dissertation and learning book, a shift became visible from simple prompt steering to process and production architecture; from trust in coherent output to explicit verification; from a single generative system to organised counterplay between different systems; and from human text production to human curation. This also made clear that the quality of AI-generated academic production depends not only on the capabilities of the language model used, but also on how generation, critique, verification and human decision-making are organisationally separated.

In retrospect, the study therefore acquired a role that had not been anticipated at its inception. It originated before the pentalogy *AI-driven Knowledge Production,* but became its historical first case and represents the design-oriented archetype within it. Several practices that could be specified in advance in the later four studies still had to be developed during the process here: the curatorial role of the human researcher, the use of AI as a text producer rather than solely as a writing assistant, the interchangeability of production and critique roles between models, adversarial control, verification of sources and claims, and the transition from prompting to

reproducible process architecture. These experiences yielded methodological principles that could subsequently be specified and applied more systematically in the later studies.

Once the pentalogy has been completed, this first case acquires renewed significance. What still had to be developed during the process here can then be understood as part of a broader methodological framework for AI-driven knowledge production. The question thereby shifts from whether full AI text production under human curation is possible to what this division of labour means when it is systematically applied across diverse scientific and societal domains of knowledge. The *Unvollendete* thus stands chronologically at the beginning of the pentalogy and has also provided important methodological principles for what followed.

# 2. Management Summary: Governing within Biological Boundaries, The Role of Biological Constraints in Addressing Persistent Problems in Public Administration

## Chapter 1 – Problem Definition and Research Context

The first chapter situates the study within the broader public administration literature. Explanations of policy failure have traditionally focused on economic, legal, and religious-cultural factors. These models assume that citizens rationally maximise utility, comply with rules, or subscribe to shared values. In practice, however, policies that appear internally coherent when designed along these lines often encounter unexpected and persistent resistance. Examples from Dutch public administration show that interventions which appear rationally persuasive regularly falter during implementation.

At the heart of the problem lies a missing dimension: human nature. Biological mechanisms such as status defence, territorial attachment, and group loyalty influence behaviour in ways that traditional models do not fully explain. This dissertation therefore starts from the observation that these evolutionarily shaped mechanisms may help determine policy outcomes. The resulting central question is: how can biological constraints function as an integrating factor within public administration, and what does this mean for our understanding of policy outcomes?

## Chapter 2 – Theoretical Foundations

This chapter further develops the central research question — '*How can biological constraints function as an integrating factor within effective public administration, and what practical implications does this have for policy development and implementation?*' — by drawing on insights from evolutionary psychology, anthropology, and the behavioural sciences. These disciplines suggest that much human behaviour may be traced to mechanisms shaped during human evolution. Examples include tendencies towards hierarchy formation, territorial defence, loyalty to one's own group, and parents' propensity to invest in their children.

On the basis of this literature, a framework of ten biological constraints with potential relevance to policy is developed. These constraints are not conceived as rigid laws, but as structural conditions that influence the circumstances under which policy interventions are likely to succeed. The chapter also introduces the concept of evolutionary-biological robustness: policy is robust when it takes account of these fundamental behavioural mechanisms and may fail when they are disregarded or suppressed.

## Chapter 3 – Theoretical Research Question 1

The first theoretical research question is: '*Which fundamental types of conflict are distinguished within public administration, how can they be classified according to biological patterns, and*

*why are certain conflicts persistently resistant to traditional forms of regulation?*' The literature review identifies ten mechanisms that may be understood as constraints. These comprise carrying-capacity constraints, resource competition, reproductive strategies and mate selection, territorial attachment, social hierarchy, altruism and reciprocity, parental investment, frequency dependence and environmental variability, life-history strategies, and coalition formation.

The chapter discusses the theoretical foundations of each of these constraints and the ways in which they may be identified empirically. Together, they constitute a framework that may not only deepen public administration scholarship, but also provide a basis for examining concrete policy cases.

## Chapter 4 – Theoretical Research Question 2

The second theoretical research question is: '*Under what conditions do the anthropological assumptions underlying economic, legal, and religious regulatory mechanisms undermine their own effectiveness, and what role do biological constraints play in this process?*' This chapter presents ten evolutionary-biological mechanisms and uses empirical illustrations to show how they may facilitate or impede policy implementation.

It further argues that classifying and interpreting public administration problems in terms of class conflict, racial conflict, and conflict between the sexes may offer a productive analytical perspective. Although this tripartite classification is not absolute, its potential strength lies in distinguishing three arenas in which biological mechanisms may manifest themselves in different ways. Class conflict primarily concerns status and investment in children; racial conflict concerns territory and group loyalty; and conflict between the sexes concerns reproductive strategies and power relations. At the same time, the mechanisms are recognised as extending across domains: status plays a role in all three arenas, but takes a different form in each.

## Chapter 5 – Theoretical Research Question 3

The third theoretical research question is: *Which biological constraints are most relevant to public administration, how can they be integrated into policy theory, and in what ways can they complement traditional regulatory mechanisms without replacing them?* This chapter develops a conceptual framework that enables biological insights to be integrated into policy theory. It formulates three theoretical conditions for such integration: structural heterogeneity, epistemic boundedness, and functional complementarity. Together, these conditions determine the circumstances under which evolutionarily informed explanations can be meaningful within public administration. The analysis shows that biological insights should not serve as alternatives to existing models, but as a plausibility filter through which economic, legal, and cultural explanations can be more effectively aligned. This produces a policy theory that does not regard human behaviour as infinitely malleable, but as embedded within biological and social constraints.

## Chapter 6 – Research Design and Methodology

This chapter translates the theoretical conditions developed in Chapter 5 into a diagnostic analytical framework for the empirical study of policy processes. Six cases, distributed across the three fields of conflict — class conflict, race conflict, and sex conflict — were purposively

selected because of their contentious and persistently intractable character. The approach is qualitative and abductive, and focuses on identifying patterns of policy failure associated with the neglect of biological constraints. Each case is reconstructed chronologically, after which the cases are compared in order to identify recurring mechanisms. The analysis uses pattern matching and the triangulation of documents, policy statistics, and discourse analysis. Validity is analytical in nature: the value of the study lies in the plausibility of the patterns identified, rather than in statistical generalisation.

## Chapter 7 – Class Conflict

This chapter analyses two cases within the domain of class conflict: executive remuneration and bonus regulation, on the one hand, and educational equality and parental privilege, on the other. Both cases examine why redistributive policies and meritocratic systems encounter structural resistance despite the widely shared rhetoric of equality and justice.

The first case shows how executive remuneration continued to escalate despite economic theories of optimal incentives, statutory bonus caps, and public indignation. Status competition is the central explanatory mechanism: remuneration committees systematically benchmark above peer-group averages — the so-called ratcheting effect — causing remuneration levels to rise irrespective of individual performance. Coalition formation within elite networks reinforces this pattern: non-executive directors and executives exchange roles within these networks, allowing reciprocal status affirmation to take the place of economic rationalisation. Statutory caps and say-on-pay regulation had limited effects because they did not address the underlying status logic, but merely sought to constrain its outcomes. Biologically informed alternatives that channel status competition through collective reputational mechanisms — as illustrated by Swiss and Japanese remuneration models — display greater durability.

The education case shows how equal-opportunity policies encounter parental investment as a biological mechanism. Parents systematically engage in behaviour that maximises the prospects of their own children, even where this undermines collective objectives of equality. Early-selection systems such as the Dutch system reinforce this mechanism by concentrating parental influence at an early and decisive point. Finnish and Swiss education systems, which postpone definitive selection or distribute status across multiple educational routes, appear to limit some of the opportunities for this dynamic to operate and are associated with smaller socio-economic gradients in educational outcomes.

## Chapter 8 – Race Conflict

This chapter analyses two cases within the domain of identity conflict: immigration and social cohesion (Case A), and educational segregation (Case B). Both cases examine why policies intended to promote ethnic integration display systematically limited effectiveness, and what role territorial attachment and coalition formation play in this outcome.

The immigration case compares five countries — the Netherlands, Canada, France, Germany, and Sweden — which followed markedly different integration trajectories despite having broadly comparable resources and objectives. Traditional instruments — economic investment in disadvantaged neighbourhoods, legal anti-discrimination frameworks, and cultural awareness campaigns — had limited effects in all five countries on actual social cohesion and residential segregation. The biological analysis identifies three mechanisms that explain this

pattern. Territorial attachment generates predictable behavioural responses once the concentration of residents with a migration background exceeds a certain threshold, irrespective of the objective liveability of the area. Coalition formation produces ethnic clustering that is consistent across national contexts, regardless of whether official policy is based on multiculturalism or *républicanisme*. Perceptions of carrying capacity respond more strongly to the pace of demographic change than to absolute numbers, helping to explain why periods of rapid inflow generate disproportionate public unease.

The crucial difference between countries lies not in whether these mechanisms are activated — they are treated here as universal — but in the institutional response to them. Canada's proactive accommodation through a points-based system and geographical dispersal policies prevented demographic shocks capable of activating territorial defensive responses. France and Germany opted for the institutional denial of group identities, causing clustering to become less visible and more difficult to channel.

The educational segregation case analyses three Dutch cities — Amsterdam, The Hague, and Rotterdam — which implemented different dispersal measures between 2010 and 2021. Economic measures to improve the quality of so-called 'black schools', legal mechanisms involving lotteries and postcode-based allocation, and cultural awareness programmes all had minimal effects on the ethnic composition of schools. Parents engaged in systematic compensatory behaviour — including increased use of denominational and other privately governed schools, residential mobility, and parallel educational pathways — which neutralised the effects of enforced mixing.

Residential segregation influences the local pool of pupils and therefore constitutes an important condition for school segregation, but territorial attachment does not independently explain actual school-choice behaviour. Coalition formation helps explain preferences for a socially recognisable school environment, frequency dependence explains how successive choices may reinforce changes in school composition, and parental investment helps explain why parents accept substantial costs when they perceive risks to their children.

The biologically informed alternatives derived from the case — phased changes in school composition, the development of a superordinate school identity, and adaptive implementation accompanied by monitoring — should be regarded as testable design options rather than as proven solutions. The central conclusion is that integration policy must address the institutional conditions under which segregation is reproduced, rather than treating individual school choices in isolation.

## Chapter 9 – Sex Conflict

Chapter 9 analyses two cases within the domain of sex conflict: the gender pay gap in the Netherlands (Case A) and the issues associated with #MeToo (Case B). Both cases examine why gender inequality and asymmetries of power persist despite decades of legal, economic, and cultural intervention.

The gender pay gap case compares three countries — the Netherlands, Iceland, and the United Kingdom — each of which adopted a different policy route. The Netherlands implemented EU Directive 2023/970 through pay transparency without substantial sanctions. Iceland operates a

certification system that actively enforces equality by placing the burden of proof on the employer. The United Kingdom introduced mandatory disclosure through Gender Pay Gap Reporting. Despite this variation, policy outcomes converged: pay differentials persist in all three countries.

The biological analysis identifies three mechanisms. Reproductive strategies and differentiated preferences explain systematic differences in risk tolerance and preferences for stability that structure choices of sector and contractual arrangement. Male norms of status competition have shaped organisational cultures that value input over output and treat continuous availability as the standard — criteria that systematically favour employees without substantial caring responsibilities. Environmental sanctions discourage assertive behaviour by women through social and professional penalties that men displaying equivalent behaviour do not encounter. Policies that merely increase transparency without addressing these three mechanisms fail to engage with the institutional architecture that sustains pay differentials. Biologically informed alternatives — output-based job evaluation, nuanced transparency that acknowledges legitimate variability, and the diversification of status through multiple dimensions of recognition — have demonstrated greater effectiveness in limited trials.

The #MeToo case analyses four countries — the Netherlands, France, Sweden, and Canada — which display systematically limited policy effectiveness despite broadly shared objectives. Legal procedures encounter structural evidential difficulties. Economic sanctions are too weak to compel organisations to adopt preventive measures. Cultural interventions achieve changes in attitudes without corresponding changes in behaviour.

The biological analysis identifies three mechanisms that operate primarily within the surrounding environment, rather than at the level of victims. Social hierarchy and status competition facilitate harassment where power is highly concentrated and the expected costs of sanctions remain low: 73 per cent of reported cases involve a perpetrator in a higher hierarchical position. Perceptual asymmetry, understood as a manifestation of reproductive strategies, generates a category of transgressions in which the perpetrator is unaware that the conduct crosses a boundary, whereas a substantial proportion of other conduct is conscious and intentional — a distinction with direct policy implications. Coalition formation protects accused individuals institutionally through implicit in-group loyalty which, in male-dominated organisations, automatically prioritises protection of the group's own members over the proper handling of complaints.

Biologically informed alternatives — the rotation of power and external oversight, explicit consent frameworks, the recording of warning signals and detection of recurring patterns, and gender-diverse decision-making teams — each address a specific mechanism at the structural level. The central conclusion is that effective policy intervenes in the architecture of power structures and coalition dynamics, rather than focusing on individual attitudes or intentions.

## Chapter 10 – Synthesis

This chapter answers the central research question. The six cases suggest that the same biological constraints recur across different policy domains. Sensitivity to status, territorial attachment, coalition formation, and parental investment were found to exert a significant influence on policy outcomes.

The chapter also clarifies why the three traditional policy instruments may display limitations. Economic instruments assume that people rationally maximise utility, but may overlook the fact that relative status can carry greater weight than absolute prosperity. Legal instruments assume that rules will be followed, but may come into conflict with group loyalty and territorial boundaries. Religious-cultural instruments assume shared values, but may overestimate the willingness to include other identities where exclusionary mechanisms predominate.

Biological constraints may therefore function as conditions that help determine when traditional instruments are or are not effective. The framework of ten constraints appears to be largely valid: some mechanisms probably operate structurally, whereas others are context-dependent or less readily observable. The three-conflict taxonomy has proved useful as an organising device, although the mechanisms frequently overlap across domains.

The conclusion is that evolutionary-biological constraints may constitute an integrating factor within public administration. They may help explain why policy fails or succeeds, depending on the extent to which they are disregarded or acknowledged. They are, however, descriptive rather than prescriptive, unlike legal, economic, or religious-cultural policy instruments, and they are certainly not deterministic: they may help explain the failure of traditional policy instruments, but such explanations remain context-dependent.

## Chapter 11 – Practical Application

The final chapter translates the findings into policy practice. Here, evolutionary-biological robustness is developed as a possible criterion for assessing policy. A policy may be considered robust when it takes account of biological constraints that may be deeply embedded in human behaviour.

The EBR scan is introduced as a brief, early-stage assessment in which policymakers ask five core questions: which status positions are affected, which territorial sensitivities are at stake, which group boundaries are being redrawn, what consequences arise for children and future generations, and whether the policy affects differing preferences and priorities between the sexes, life stages, or partner dynamics. The EBR assessment develops this process further by evaluating interventions systematically against the ten constraints, formulated as Ten Commandments. The purpose is not rigid application, but to increase the likelihood that policy will prove socially sustainable.

## Concluding Reflections

The dissertation demonstrates that policy should not be understood solely in legal, economic, or cultural terms, and that evolutionary-biological constraints may constitute a crucial analytical dimension. By incorporating these constraints systematically into the analysis, a more realistic and durable understanding of effective governing may be developed. The study's contribution lies in making visible a dimension that often remains implicit: the possible role of human nature. At the same time, the conclusions are explicitly qualified, and further research is recommended into the relative hierarchy of the constraints, their context-dependence, and the dynamics of normative change.

# 3. Management Summary - AI-Driven Knowledge Production- A Study of the Impact of Artificial Intelligence on Scientific Knowledge Production

This dissertation forms the concluding part of a pentalogy of five academic monographs. The four preceding monographs described the use of AI in the natural sciences, history, anthropology and design-oriented research. In each case, AI was used to write a scientific dissertation in accordance with the conventions of the relevant domain. The present dissertation is a meta-dissertation: it serves as an umbrella over the other monographs. It organises science by identifying four-plus-one archetypes of scientific practice and subsequently examines how the use of AI affects these forms of scientific practice. The four aforementioned monographs therefore constitute the case material for this dissertation.

## Chapter 1 – Introduction: AI and Knowledge Production

This dissertation centres on the fundamental question of how artificial intelligence (AI) is changing the foundations of scientific knowledge production. Whereas computers previously functioned primarily as calculating tools, they have now become co-producers of knowledge. This development raises urgent questions concerning the ways in which AI affects processes of data collection, analysis, explanation and presentation.

To structure this complex subject matter, the chapter introduces four archetypes of knowledge production. The first archetype concerns natural science research, which seeks to discover universal laws through measurement and experimentation. The second encompasses historical research, in which the understanding of unique events is central and is pursued through source criticism and interpretation. The third comprises anthropological research, in which culture is understood through participant observation and thick description. The fourth concerns design-oriented research, in which knowledge is produced by creating solutions.

Alongside these four archetypes, an important distinction is made between generic AI — large models trained on internet data — and domain-specific AI, trained on carefully selected datasets. In addition, the meta-domain of computer science and AI development itself receives special attention because of its reflexive character, with AI simultaneously serving as an instrument and an object of research.

The analytical framework used in this dissertation comprises six impact dimensions: cognitive, methodological, pedagogical, institutional, power and ownership, and reliability and normativity. Particular attention is paid to what is termed the AI paradox: the paradoxical situation in which AI can be used productively only by those who have first been educated without it.

## Chapter 2 – Research Questions

The second chapter formulates the central research questions against the background of what is described as fragmentation within contemporary academic culture. Although AI accelerates research processes, it simultaneously risks reinforcing the regime of 'postage-stamp insights' — small, narrowly delimited contributions — rather than promoting genuine synthesis.

The main research question is: How does the use of artificial intelligence affect the processes, institutions and epistemology of scientific knowledge production, and what scope does this create for overcoming fragmentation and promoting multidisciplinary originality?

This main question is elaborated through three sub-questions. The first examines how AI affects the cognitive and methodological dimension, specifically the ways in which knowledge is sought, organised and verified. The second addresses the pedagogical and institutional consequences, focusing on the impact on educational practices and scientific recognition. The third examines the dimension of power and normativity, with particular attention to how AI affects the distribution of power and ownership within science.

The scientific relevance of this research lies in its provision of an integrated perspective that overcomes the fragmentation of the existing literature on AI. From a societal perspective, it is relevant because AI affects both the reliability of knowledge and the education of new generations.

## Chapter 3 – Methodological Rationale

This chapter explains how the research was designed. Because AI is both the subject of study and the instrument used in conducting the research, a distinctive methodological situation arises that requires a hybrid approach.

The research combines several methods: literature research, analysis of existing AI applications across different scientific domains, and reflection on the researcher's own experiences of using AI. This combination is necessary because AI is still so new that few established research traditions exist.

The research structure is organised as a systematic comparison. Four different types of science — the natural sciences, history, anthropology and design-oriented research — are each examined in relation to three aspects: how AI changes thinking and working practices, how it affects education and institutions, and which power relations emerge. This systematic approach prevents the research from becoming lost in disconnected observations. For each scientific domain, existing literature on AI applications is used and supplemented with practical examples. In addition, three concrete cases are examined in which the researcher worked intensively with AI, in order to gain insight into how AI-driven knowledge production operates in practice.

Quality assurance is achieved by comparing different sources and perspectives, transparently documenting every step of the research, and continuously reflecting critically on the limitations of both AI and the human researcher.

## Chapter 4 – Cognitive and Methodological Impact

The fourth chapter analyses how AI reconfigures the fundamental processes of knowledge production. One important shift identified concerns the changing role of researchers: from data collectors to curators of AI-generated insights. Although AI enables large-scale replication, a problematic shift is also occurring: AI can often predict very accurately what will happen — such as which patient will become ill or which protein structure will emerge — while scientists increasingly understand less about why that prediction is correct. This undermines the traditional scientific ideal of knowing not only what happens, but also why.

In terms of opportunities, AI is found to enable considerable acceleration of literature research and data processing. It also makes feasible large-scale reanalyses and replication studies that were previously inconceivable. AI creates cognitive space for interpretation by automating repetitive tasks and enables pattern recognition across disciplinary boundaries.

These opportunities are accompanied, however, by substantial risks. These include oversimplification and loss of nuance, as well as a particularly problematic phenomenon: the fabrication of bibliographic references, which may range from accurate to entirely fictitious. There is also the reproduction of biases contained in training data, while the problem of black-box opacity remains.

The evaluation of these developments shows that the tension between acceleration and depth is crucial to the future of knowledge production. AI may reinforce fragmentation by enabling the production of more papers, but it can simultaneously promote synthesis by revealing patterns. The multiplication of AI reviewers may nevertheless create echo chambers if all models have been trained on similar data.

## Chapter 5 – Pedagogical and Institutional Impact

The fifth chapter examines how AI changes both learning and the institutional structures of science. A fundamental observation is that the writing process is becoming a learning environment, while traditional rituals such as dissertations and peer review are coming under pressure.

Several pedagogical opportunities are identified. AI enables the personalisation of learning and improves accessibility for students with disabilities. It can also increase the transparency of learning processes through automatic logs that document the learning trajectory.

Institutional shifts manifest themselves at several levels. Publication culture is accelerating, but may also become more superficial. Peer review can be automated, but risks losing its human dimension. In addition, the criteria governing academic careers must be recalibrated to reflect new forms of knowledge production.

Three fundamental tensions are identified that affect the core of these changes. The first concerns the communicative function of scientific texts: they are not merely vehicles for

information, but also rituals that create community. The second involves the AI paradox: students who grow up with AI lack the frame of reference needed for critical evaluation. The third concerns the tension between originality and fragmentation: AI can both reinforce and overcome postage-stamp insights.

Although the Hong Kong Principles — integrity, transparency and open science — provide frameworks for these developments, new rituals for an AI era have yet to be developed.

## Chapter 6 – Power, Normativity and the Nature of AI-Driven Knowledge Production

The sixth chapter focuses on a crucial question: who actually determines what counts as reliable knowledge when AI is used? The answer proves to lie not only with scientists, but also with the companies and institutions that develop and own AI systems.

Power and access play a central role because training AI is extremely expensive. Only large technology companies can afford systems such as GPT or AlphaFold. This means that smaller universities and researchers become dependent on what these companies make available. As a result, researchers do not have equal opportunities to produce knowledge.

Embedded choices within AI systems help to determine what counts as knowledge. When an AI model is trained, decisions are made about which data are included and which are excluded. A model trained predominantly on English-language texts will identify different patterns from one trained on a diversity of languages. These choices are made by developers rather than by the scientists who later use the model.

Data function as an instrument of power because high-quality datasets are scarce. Medical records, historical archives and social-media data are not equally accessible to everyone. Those who control these data partly determine which research is possible and which questions can be asked.

The black-box problem means that researchers often do not know how an AI system reaches its conclusions. Even in specialised systems such as AlphaFold, it is not always clear why the model produces a particular prediction. This makes it difficult to verify results or detect errors.

The chapter concludes that the use of AI in science is never neutral, but is always influenced by who has access to technology and data, and by the choices embedded within the systems.

## Interim Assessment 1 – Taxonomies of Opportunities, Risks and Challenges

Following the treatment of the three dimension chapters, the insights are brought together in systematic taxonomies that serve as the analytical framework for examining the archetypes.

The taxonomy of opportunities comprises three main categories. The first is acceleration and scale, allowing feedback loops to be radically shortened and processes to become cyclical. The second concerns discovery and synthesis, through which relationships across disciplinary boundaries become visible. The third comprises pedagogical and institutional innovation through personalised learning and new forms of transparency.

The taxonomy of risks identifies five core problems. The first concerns black boxes and loss of transparency, which undermine reproducibility. The second comprises decontextualisation and loss of nuance, creating an appearance of universality around contextually contingent patterns. The third concerns ritualisation and loss of meaning, resulting in the erosion of science's community-forming function. The fourth involves the concentration of power, with epistemic agendas being shaped by commercial interests. The fifth concerns the fabrication of references, which threatens source attribution.

The taxonomy of challenges identifies three central problems: how can we determine whether AI-supported research is reliable when we cannot inspect the underlying processes? How can the meaningful rituals of science, such as dissertations and peer review, be preserved while AI is integrated? And how should different disciplines establish their own standards for the degree of AI-related uncertainty that is acceptable?

## Chapter 7 – Natural Science Research and AI

The seventh chapter analyses how AI affects the first archetype. The natural sciences, traditionally centred on reproducible measurements and falsifiable hypotheses, exhibit the most harmonious relationship with AI because their value systems converge around empirical verification.

The opportunities AI offers natural science research are substantial. It can accelerate discovery processes in domains such as molecular biology and astronomy. The analysis of large datasets creates new possibilities in climatology and other data-intensive fields. The democratisation of knowledge through public databases improves accessibility, while advanced simulations and modelling increase methodological precision.

These opportunities are nevertheless accompanied by risks. The problem of black-box explainability persists even when predictions are accurate. Dependency on closed infrastructures is a growing concern. Bias in training data may create a misleading appearance of objectivity. High energy and resource costs also affect accessibility.

The challenges manifest themselves at three levels. Methodologically, the question arises whether accurate but inexplicable predictions are scientifically valid when the underlying mechanism is not understood. Institutionally, smaller research groups must also be given access

to expensive infrastructure. Normatively, each domain must determine how much uncertainty is acceptable.

The conclusion is that AI functions primarily as an accelerator in the natural sciences, reinforcing existing values while also introducing new dependencies and normative choices that must be addressed explicitly.

## Chapter 8 – Historical Research and AI

In historical scholarship, where authenticity and context are central, AI creates specific tensions concerning anachronism and canon formation. This tension forms the central thread of the analysis.

The opportunities offered by AI are considerable. Access to digital archives is greatly improved, while pattern recognition can reveal unexpected relationships. OCR and translation tools substantially increase the accessibility of sources. Education and institutional innovation become possible through the more active role archives can assume in knowledge production.

These opportunities are accompanied by significant risks. Anachronistic interpretations may arise when contemporary categories are applied to historical data. The datafication of sources threatens authenticity and context. Black-box patterns may be adopted uncritically. Canon formation through selective digitisation presents a particular problem.

The challenges centre on three key issues. Methodologically, historians must always translate AI-generated patterns back into the historical contexts in which the sources originated. Institutionally, decisions made during digitisation must be made explicit because they determine which narratives become visible. Normatively, the historian remains responsible for judging whether AI-supported analyses are historically valid.

The conclusion is that historical scholarship can use AI to improve access and pattern recognition, but that human interpretation remains essential to prevent anachronism and safeguard authenticity.

## Chapter 9 – Anthropological Research and AI

The analysis of the third archetype shows that anthropology, founded on relational proximity and contextual richness, experiences the greatest friction with AI's tendency towards datafication and pattern recognition. This fundamental tension shapes the entire treatment of this archetype.

The opportunities offered by AI for anthropological research include several innovative applications. Digital ethnography enables analysis on a larger scale, while AI-assisted transcription and thematic coding can accelerate the research process. The documentation of endangered languages and rituals becomes technically feasible, while remote sensing creates new possibilities for landscape analysis and the study of mobility patterns.

The associated risks are substantial, however. The reduction of communities to data points threatens the core of anthropological work, while the loss of proximity and empathy undermines its methodological foundations. AI may impose classifications that do not correspond to how communities understand themselves — for example, by labelling ritual gestures as everyday behaviour or identifying universal patterns where local meanings differ considerably.

The challenges facing anthropological research affect the foundations of the discipline. Methodologically, AI must function as a supplement to, rather than a replacement for, interpretative work. Institutionally, it is essential that data be returned to research participants in order to preserve relational legitimacy. Normatively, consent must be guaranteed relationally and continuously.

The conclusion is that anthropology must explicitly embed AI within the values of proximity, empathy and shared responsibility in order to preserve its disciplinary integrity.

## Chapter 10 – Design-Oriented Research and AI

Within this fourth archetype, AI acts as a co-designer, creating new possibilities while simultaneously raising fundamental questions concerning the explainability of artefacts and responsibility for them.

The opportunities offered by design-oriented research are impressive. AI as a co-designer can generate thousands of design variants, thereby broadening the creative horizon. Faster and less expensive simulation and prototyping make iterative development more accessible. Iterative cycles are greatly shortened, increasing efficiency. New forms of pedagogy become possible through direct collaboration between students and AI tools.

These promises are nevertheless accompanied by risks. Black-box designs are difficult to explain, which undermines the legitimacy of design choices. Accountability for design decisions may be lost when AI makes autonomous decisions. Ownership and authorship become diffuse within hybrid human–machine processes. Moreover, AI-supported optimisation may disregard moral values that have not been explicitly programmed.

The challenges facing design-oriented research centre on three core areas. Methodologically, standards of explainability must be developed to safeguard transparency. Institutionally, new forms of publication and recognition are needed that do justice to hybrid authorship. Normatively, values and ethics must be explicitly embedded in design processes.

The analysis concludes that AI can strengthen design-oriented research, provided that explainability, institutional recognition and collective accountability are systematically incorporated.

## Chapter 11 – The Meta-Domain of AI & Computer Science

The meta-domain is distinctive because AI is both the instrument and the object of research — research into AI-driven knowledge development conducted using AI — which creates reflexivity and circular validation. This distinctive position makes the domain an instructive test case for the broader analysis.

The utopian scenario presents impressive possibilities. AutoML and meta-learning create self-improving systems capable of exponential development. Open platforms such as arXiv and Wikipedia promote collective ownership and democratise knowledge production. New pedagogical forms become possible in which learning about AI takes place through AI, thereby developing reflexive competencies.

The dystopian scenario, however, warns of serious dangers. Circular validation against self-defined benchmarks may render apparent progress illusory. Extreme concentrations of power among a small number of companies threaten to undermine research autonomy. The black-box character becomes so pronounced that even designers no longer understand their own systems. The fragmentation of legitimacy across different validation practices undermines scientific consensus.

The challenges facing the meta-domain are wide-ranging. Methodologically, criteria must be developed that break circular validation and safeguard external validity. Institutionally, a balance must be found between open and closed systems. Normatively, responsibility must be redefined within a recursive ecosystem in which traditional boundaries become blurred.

Important qualifications are provided by counterexamples such as Wikipedia and arXiv, which demonstrate that alternative forms of validation are possible. Older research traditions such as hermeneutics, autoethnography and Actor–Network Theory offer conceptual resources for reflexive legitimation that remain relevant in the AI era. These traditions teach, respectively, that AI results always require interpretation through hermeneutics, that self-study can be legitimate provided it is accompanied by transparent reflection through autoethnography, and that knowledge always emerges within networks of people and technologies through Actor–Network Theory.

## Chapter 12 – Interim Assessment 2: Overarching Patterns and Comparison of Archetypes

Chapter 12 presents a second interim assessment. It brings together the insights from the three impact dimensions and the four archetypes of science. Comparison of these domains shows that AI is applied differently across disciplines, while several recurring tensions simultaneously become apparent. The analysis demonstrates that these tensions can be understood as three fundamental paradoxes of AI-driven knowledge production.

The first paradox concerns the relationship between acceleration and legitimacy. In virtually all the domains examined, AI increases the speed and scale of analysis. At the same time, this acceleration places under pressure precisely those mechanisms that have traditionally safeguarded scientific legitimacy. In the natural sciences, for example, AI may generate

hypotheses that can be validated experimentally only at a later stage; in historical research, large-scale textual analyses may suggest patterns that are historically anachronistic; and in design-oriented research, generative AI may accelerate the creative process without the underlying design choices remaining fully transparent.

The second paradox concerns access and inequality. In many cases, AI lowers barriers to research because digital archives, analytical tools and generative systems are widely available. At the same time, new forms of dependency emerge in relation to technological infrastructure, specialised expertise and representative datasets. Existing inequalities therefore do not disappear, but are reproduced in new forms.

The third paradox concerns knowledge formation and dependency. Effective use of AI presupposes precisely the analytical skills and methodological discipline that AI systems partly take over. When researchers do not develop these skills independently, there is a risk that AI systems will be used productively while their outputs cannot be assessed with sufficient critical rigour.

This interim assessment therefore shows that AI not only creates new possibilities for knowledge production, but also introduces structural tensions. These paradoxes do not appear to be temporary transitional problems, but structural characteristics of a research practice in which human and artificial intelligence are becoming increasingly intertwined. Chapters 13 to 17 examine whether empirical experience leads to the revision or further development of these findings.

## Chapter 13 – Case Study: 'Avian Intelligence & Homing Pigeons'

This chapter analyses a natural science-style study of navigational behaviour in homing pigeons. The manuscript follows the classical architecture of an empirical dissertation, comprising a literature review, a methodological rationale, multiple datasets and a synthesis chapter in which the results are interpreted. The study concludes that pigeon navigation cannot be explained by a single mechanism, but instead involves a combination of overlapping orientation strategies.

The case shows that generative AI can play a role in different phases of the research process. Its greatest added value appears to lie in the early stages of research, such as literature exploration, hypothesis formation and the systematic examination of possible explanatory models. AI can rapidly synthesise large bodies of literature, generate alternative hypotheses and assist in structuring research designs. This can considerably accelerate the processes of problem exploration and conceptual delimitation.

At the same time, the case demonstrates that epistemic responsibility for the ultimate research architecture remains with the human researcher. The selection of hypotheses, the interpretation of results and the integration of different explanatory strands continue to require curatorial oversight and methodological discipline. In this configuration, AI functions primarily as an instrument for analysis and synthesis, rather than as an autonomous producer of scientific knowledge.

*Scientific illusionism.* The final section of the chapter, however, makes one crucial fact explicit: the experiment on which the dissertation is based never actually took place. The datasets underpinning the analysis were generated synthetically using generative AI. The case therefore functions as a controlled experiment in scientific illusionism. It demonstrates that generative AI is capable of producing an apparently empirical study that appears methodologically plausible and stylistically conforms to academic conventions. The manuscript's persuasiveness rests not on spectacular claims, but precisely on methodological restraint, plausible variation within the datasets and an interpretative style that explicitly acknowledges uncertainties and alternative explanations.

This exercise reveals that the form of scientific argumentation can itself be an important source of persuasiveness. When datasets appear plausible and the argumentation conforms to established disciplinary traditions, a study may seem convincing even when its empirical basis is synthetic. The case thereby underscores the importance of transparency concerning the provenance of data and of methodological control mechanisms that extend beyond stylistic or rhetorical plausibility.

## Chapter 14 – Case Study: 'Predictions of Digital Transformations'

This chapter analyses the production of a historical study of predictions concerning digital technology and their actual societal outcomes in the Netherlands between 1945 and 2025. The study reconstructs a canon of ICT developments presented in political and societal sources as potentially transformative. Using this canon, historical claims about the future are systematically compared with the actual societal developments that followed. To make this comparison analytically manageable, a 2×2 matrix is developed that combines two dimensions: the accuracy of the original prediction and the extent to which the predicted societal transformation was actually realised. This produces a typology of historical cases in which expectations and outcomes exhibit different relationships.

Generative AI played a role primarily in the exploratory and analytical phases of the research process. Language models were used to identify possible episodes for inclusion in the canon, to organise large quantities of historical and policy material, and to explore different classifications of technological claims about the future. AI was also used in the systematic application of three analytical lenses: a claim-oriented lens, a transformation lens and a causal lens examining the relationship between the two.

The case demonstrates that AI can support historical research by structuring and analysing extensive discursive corpora. At the same time, the assessment of historical adequacy remains dependent on human contextualisation and interpretative expertise. The analysis therefore illustrates that AI adds value primarily in organising and comparing historical material, while the ultimate historiographical interpretation remains a human responsibility.

## Chapter 15 – Case Study: ‘Governing Within Biological Boundaries’

This chapter examines the production of a dissertation describing the influence of evolutionarily and biologically shaped factors on persistent problems within public administration. The study begins from the proposition that explanations of policy failure remain incomplete when they rely exclusively on economic, legal or religious-cultural models. A framework of ten evolutionary-biological mechanisms was therefore developed. Disregarding these mechanisms in policymaking could explain the persistence of certain problems in public administration. Empirically, the study is structured around six cases distributed across three fields of conflict and is methodologically elaborated as qualitative, abductive pattern-matching research.

Generative AI played a broad role in the early and analytical phases of the research process. AI was used for literature exploration, conceptual organisation, structuring the theoretical framework and developing interpretative analyses within an interdisciplinary corpus. The case thereby demonstrates that AI adds particular value where large quantities of dispersed literature must be brought together within a more coherent analytical framework.

At the same time, the chapter exposes a classical epistemic vulnerability of interpretative disciplines. Research involving meaning-making and narrative reconstruction is traditionally susceptible to selective interpretation, narrative coherence and confirmation bias. Generative AI can reinforce these vulnerabilities because language models optimise for plausible and consistent text and can readily elaborate implicit assumptions into an increasingly coherent narrative.

It was precisely this vulnerability that led to a methodological innovation in this case. Rather than using AI solely in a generative capacity, a second model was deployed in an adversarial and supervisory role. One model generated textual proposals and analyses, while another was used systematically to identify inconsistencies, weaknesses and alternative interpretations. This created a form of role-separated modelling in which different epistemic functions were deliberately kept distinct. The case thus demonstrates that AI can not only intensify risks within interpretative disciplines, but also enable new forms of methodological counter-checking.

## Chapter 16 – Case Study: ‘Polyglot Language Acquisition’

This chapter discusses the production of a design-oriented dissertation on polyglot language acquisition. The starting point of the study is the observation that language education is traditionally organised separately for each language, whereas many multilingual speakers make use of structural similarities between languages. The dissertation examines how these similarities can be used systematically within a learning architecture for the parallel acquisition of multiple languages. The study combines insights from linguistics, cognitive science and pedagogy and develops a design framework in which grammatical structures, word formation and syntactic patterns across languages are explicitly connected. Generative AI was used to develop and refine examples, exercises and illustrations of such structures.

The case shows that AI is particularly valuable in design-oriented research for elaborating and operationalising a conceptual design. Whereas traditional design processes are often constrained by the time required to develop examples and educational concepts, AI makes it

possible to translate structural language patterns rapidly into pedagogical applications. At the same time, the chapter makes clear that a generative process imposes new requirements on the conceptual stability of the design. When examples and formulations are generated on a large scale, there is a risk that terminology or pedagogical principles will gradually and subtly shift, potentially reaching the point at which the generated exercises become entirely unusable. The case makes clear that generating and broadly elaborating new concepts with AI is fundamentally different from producing artefacts with AI on a large scale and with consistency.

The case therefore demonstrates that generative AI functions in design-oriented research primarily as an instrument for concretising and elaborating an existing design structure, while the architecture of the learning process itself is designed and safeguarded by the researcher.

## Chapter 17 – Case Study: 'AI-Driven Knowledge Production' (Meta-Case)

This chapter analyses the production of the present document as a reflexive case of AI-driven knowledge production. The study thereby functions as a mise en abyme: the research object — the impact of generative AI on scientific knowledge production — becomes visible in the process through which the dissertation itself was produced, the 'nurse-within-a-nurse effect' or 'Droste effect'. The analysis therefore focuses not on the content, but on the architecture of the research process within which these texts were produced.

The case shows that generative AI can accelerate academic production considerably. The first complete version of this dissertation was generated within two days during a single extended interaction with a language model and was immediately usable as an academic working document. This exceptional turnaround time is not, however, the study's most important insight. The speed of the process appears to have been explained primarily by the way in which the research process was organised.

Two forms of discipline proved crucial. Process discipline kept the generative trajectory focused on the research objective and prevented the writing process from becoming derailed by iterative rewrites or digressions. More important still was structural discipline: the systematic construction of an analytical framework, followed by its consistent application in the case chapters and in answering the empirical research questions, all within a predesigned manuscript architecture. This architecture determined the cadence of the entire document and ensured that the interaction with the language model remained substantively stable throughout an extended dialogue.

The case also demonstrates that an explicit research architecture can break circular validation. By distinguishing different epistemic roles — such as generation, analysis and verification — model output is prevented from implicitly functioning as its own evidence. Curatorial oversight of the process thus proves to be a necessary condition for reliable AI-assisted knowledge production.

In addition to analysing the production of this dissertation, the chapter identifies several implications of AI-driven knowledge production that extend beyond the individual case. The first concerns the role of architecture: language models produce locally coherent text, but do

not independently safeguard the global argumentative structure of a manuscript. The second concerns curatorial oversight: because contextual loss and terminological drift can gradually erode the coherence of an argument, human curation is indispensable. The third concerns the use of literature: generative AI proved not only to be a source of risk, but also a tool for systematically checking the correct use of key references. The fourth concerns the organisation of the research process: separating productive and evaluative roles proved more effective than combining both functions within a single model.

The chapter also identifies a series of practical lessons from the broader research programme: regeneration behaviour that may undermine carefully curated passages, model output that is often paraphrastic rather than textually exact, users being unable to rely on a language model's reading accuracy, and limitations in context windows that lead to semantic drift. In this case — the production of the present document — these problems could be mitigated to a considerable extent through process discipline and structural discipline.

## Chapter 18 – Synthesis and Definitive Answers

This chapter brings together the findings from the five cases and compares them with the provisional conclusions from the second interim assessment in Chapter 12. An explicit delta approach is followed: for each empirical research question, the analysis examines the extent to which the case studies confirm, qualify or correct the earlier answer.

The analysis shows that generative AI changes the cognitive and methodological dimension of knowledge production. Across the different cases, AI proves capable of accelerating synthesis and analysis considerably, while also introducing new epistemic vulnerabilities. Language models produce persuasive and coherent narratives, while empirical uncertainties and alternative interpretations may recede into the background. At the same time, these systems enable new forms of verification, for example through the systematic checking of literature use and argumentative consistency.

The pedagogical and institutional analysis shows that AI changes the character of academic work. Across several cases, the researcher's role shifts from linear author to curator of a generative process. Research and writing thereby become less a linear production of text and more an iterative organisation of generation, analysis and validation.

In the normative and political dimension, the impact of AI proves strongly domain-specific. New dependencies arise primarily at the level of infrastructure and access to models, while many normative choices in the cases examined remain connected to traditional research practices.

On the basis of these findings, the central research question is answered as follows: generative AI changes scientific knowledge production by reconfiguring the architecture of the research process. Under certain conditions, this can accelerate research, broaden it interdisciplinarily and enrich it methodologically. These effects are not automatic, however; they depend heavily on how the generative process is organised.

The cases show that the quality of AI-driven knowledge production is determined not primarily by the capabilities of individual models, but by the research architecture within which they are

deployed. Explicit role separation, curatorial oversight and transparent methodological procedures prove to be crucial conditions. The chapter concludes with a discussion of the methodological safeguards employed in the research and with recommendations for further research into discipline-specific effects of AI and robust research architectures for AI-assisted science.

## Epilogue

This dissertation has examined the extent to which artificial intelligence changes knowledge production in its nature, quality and legitimacy. The five cases have shown that AI-driven knowledge production is both promising and vulnerable. It is promising because AI can accelerate processes radically, reveal patterns that would otherwise remain hidden and enable new forms of synthesis. It is vulnerable because this same power is accompanied by structural risks: black-box opacity, bias, hallucinations and the temptation of apparent plurality.

The decisive factor is consistently the human curator. Whereas AI performs the role of executor and pattern recogniser, it is the human who selects, corrects and legitimises. Without this role, the promises of AI collapse into unreliability; with it, AI can broaden the horizon of knowledge production. Transparency, methodological integrity and reasoned curation are not optional in this respect, but necessary conditions.

These findings lead to what may be understood as a ‘Socratic oath’ for the AI era. Just as physicians are bound by the principle of first doing no harm, every researcher who uses AI should swear: first, produce no false knowledge. This oath entails recognising the limits of both human and machine knowledge, transparently accounting for the choices made, and never delegating responsibility for interpretation to the technology.

The value of this oath lies in its preventive function. It prevents the power of AI from turning into intellectual arrogance and apparent efficiency from becoming more important than epistemological integrity. It thereby provides a safeguard against the erosion of scientific legitimacy.

With conscious commitment to the Socratic oath, AI-driven knowledge production can develop into a new and legitimate practice of knowing. Without such commitment, it risks becoming an instrument that undermines trust and erodes the epistemic foundations of science.

# 4. Management Summary - The Promises of Digital Revolutions: A Historical-Analytical Study of IT Predictions and Societal Transformations in the Netherlands (1945–Present).

## 1. Problem Statement and Positioning

This dissertation examines the extent to which IT-driven societal transformations are predictable. In policy arenas, academic publications and public debates, digital technology is consistently surrounded by future-oriented claims: promises of efficiency, innovation and democratisation, but also warnings about loss of control and systemic risks. Actual societal developments, however, regularly diverge from these expectations—sometimes because of overestimation, sometimes because of underestimation, and often because of shifts in scale or configuration.

The societal relevance of this study lies in the growing dependence on digital infrastructures in government, the economy and everyday life. Policy choices and investments are partly legitimised by future-oriented claims concerning technological impact. When such claims structurally lack precision or misjudge scale, this may result in institutional vulnerability or the misallocation of resources. Academically, the study addresses a gap between futures studies, the sociology of technology and innovation, and public administration. Whereas the existing literature often focuses on individual cases or abstract theorising, this dissertation systematically compares multiple waves of ICT development within a single national context.

The central research question is: 'To what extent are IT-based societal transformations predictable, and which characteristics of predictions and predictors are associated with structurally adequate predictions?' This question is elaborated through four sub-questions concerning the analysis of future-oriented claims, the nature of realised transformations, the role of technology within broader configurations, and the identification of recurring patterns of adequate prediction.

Chapter 1 accordingly positions the study as a methodological and historical-analytical investigation of the relationship between expectation and transformation in the digital domain. Its purpose is not to assess individual projects, but to reconstruct systematically the conditions under which digital future-oriented claims prove structurally adequate or inadequate.

## 2. Conceptual Framework and Theoretical Positioning

Chapter 2 develops the conceptual foundation of the study. The central problem—the predictability of IT-driven societal transformation—requires an analytical framework capable of systematically assessing both future-oriented claims and realised institutional changes. The chapter therefore positions the study at the intersection of the sociology of technology and innovation, public administration, and theories of societal modernisation.

The first step is to distinguish between technological innovation as an artefact and as a process of regime formation. Digital technology is not regarded as an isolated tool, but as part of broader socio-technical configurations in which infrastructure, regulation, organisational routines and user practices converge. In this respect, the study draws on insights from the literature on socio-technical transitions and path dependence, in which technological choices may generate cumulative institutional consequences.

The distinction between future-oriented claims and societal transformation is subsequently elaborated. Future-oriented claims may be policy-based, visionary or theory-informed, but vary in the extent to which they make causality, scale and boundary conditions explicit. Societal transformation is defined as the durable reconfiguration of institutional, organisational or socio-technical structures, rather than as temporary adoption or a passing trend.

On the basis of this conceptualisation, three analytical lenses are introduced.

The claim-oriented lens (X) analyses the quality of future-oriented claims. Predictive quality is understood as multidimensional and includes structural adequacy—whether the causal structure is explicit and plausible—conditional sensitivity—whether boundary conditions and dependencies are specified—and scale consistency—whether the assumed level of impact is coherent.

The transformation-oriented lens (Y) examines the nature of realised societal transformation. It distinguishes between sectoral optimisation, regime shifts and system-level reconfiguration, as well as between temporary changes and changes that have become durably embedded.

The causal lens (X–Y) analyses the role of technology within open, multicausal configurations. Technology is not treated as an autonomous cause, but as a potentially necessary, facilitating or constitutive factor within broader institutional and economic dynamics.

Combining these three lenses produces a systematic framework through which diverse digital episodes can be compared without losing their context-specific character. The chapter emphasises that predictability is not equivalent to exact forecasting, but concerns the extent to which articulations are structurally and configuratively adequate in relation to subsequent developments.

Chapter 2 thus provides the conceptual instruments used to analyse the historical cases in later chapters. It constitutes the methodological backbone of the dissertation: without this differentiation between claim, transformation and causal configuration, comparisons between waves of ICT development would remain analytically arbitrary.

## 3. Research Design, Analytical Method and Quality Assurance

Chapter 3 justifies the empirical approach used for the historical analysis of ICT-related future-oriented claims and societal transformations. Its central message is that the study does not consist of a series of separate narratives, but follows a standardised analytical cycle: (i) delimitation of a canon and selection of episodes; (ii) reconstruction of claims and developments on the basis of verifiable sources; (iii) application of a uniform analytical framework for each lens; and (iv) synthesis through comparable forms of output.

**Research strategy and design choices.** The chapter positions the study as qualitatively dominant and analytically comparative. Its purpose is not statistical generalisation, but the identification of recurring characteristics of adequate and inadequate prediction across different waves and outcomes of ICT development. The design is therefore variation-oriented: cases are selected to enable systematic

comparison between types of outcome, scale and regime-related subject matter, rather than to reconstruct a single ‘best case’ or ‘worst case’.

**Canon formation and selection as methodological delimitation.** In Chapter 3, canon formation functions as a form of methodological delimitation, rather than as a comprehensive historical overview or a normative ranking. A development is considered eligible only when: (1) it involves a relevant IT configuration; (2) explicit future-oriented claims or programmatic articulations are available; (3) observable societal developments can be identified; and (4) its relevance to the Netherlands can be demonstrated. This prevents the retrospective attribution of significance to episodes for which no comparable claim was articulated at the time.

**Structuring through the 2×2 matrix and avoiding success bias.** The canon is subsequently structured along two retrospective dimensions—the accuracy of the claim and the realisation of transformation—in a 2×2 matrix. The purpose is to ensure explicit variation in the empirical basis: the analysis includes not only ‘successes’ and ‘failures’, but also underestimated transformations and overestimated interventions. Success bias is thereby explicitly neutralised as a principle of selection.

**Purposive sampling and case delimitation.** A second selection is made within the canon through purposive sampling of episodes for in-depth analysis. This selection is analytically motivated by variation in scale, type of claim, outcome and source availability. Alternative cases are not allowed to disappear without explanation, but are excluded on stated grounds in order to limit implicit selection bias. Case delimitation also clearly defines the unit of analysis: a future-oriented claim, a technological cluster or an institutional configuration.

**Procedural description and reproducibility.** The chapter makes the selection and analytical process explicit through a sequence of steps: inventory → assessment → provisional classification → selection → documentation. Reproducibility is understood here in a methodological sense. Others are not expected to construct exactly the same canon, but should be able to determine why particular choices were made and which criteria were applied.

**Operationalising the ‘conceptual lenses’ as concrete analytical steps.** Crucially, Chapter 3 specifies not only what is analysed, but how the analysis is conducted. For each lens, it explains how claims are reconstructed, which dimensions are assessed and how outputs are recorded to make the cases comparable. The chapter thereby bridges the conceptual framework of Chapter 2 and its application in Chapters 5–8: each episode is examined through the same analytical grid, ensuring that the subsequent synthesis is methodologically derived rather than ad hoc.

**Quality assurance and the prevention of interpretative arbitrariness.** Finally, the chapter describes the internal quality-assurance procedures: transparent decision rules, traceability to sources and consistent output formats, including the explicit specification of claims, conditions and scale, as well as lens-specific conclusions. The central argument is that interpretative judgements are unavoidable in qualitative historical research, but can be disciplined through explicit procedures, fixed categories and verifiable substantiation.

In summary, Chapter 3 provides the procedural infrastructure of the dissertation: a verifiable route from source material to comparable outcomes for each analytical lens. This makes the case-study chapters, Chapters 5–8, mutually comparable and provides the necessary basis for the systematic pattern analysis in Chapter 9.

# 4. Canon Formation and Positioning

Chapter 4 justifies the delimitation of the empirical domain. Canon formation does not serve here as a historical overview or an assessment of technological importance, but as a methodological selection of cases suitable for the systematic comparison of future-oriented claims and societal transformation. The chapter operationalises the selection criteria developed in Chapter 3 and determines which digital interventions qualify for further analysis.

**Two gateways to inclusion.** Admission to the canon proceeds through two complementary gateways. Political articulation (P1) concerns the explicit formulation of digital ambitions or problem diagnoses in parliamentary or administrative arenas. This enables the reconstruction of original future-oriented claims. Ex post structural transformation (P2) concerns demonstrable and durable reconfiguration of institutions, markets or everyday practices, irrespective of prior political visibility. This prevents market-driven or client-side innovations from remaining outside the analytical field. A case, or 'episode', may be included through either or both gateways; neither takes precedence over the other.

**Systematic inventory.** To avoid arbitrariness, the longlist is generated through a regime × domain grid. Regime types, such as mainframe, client-side, network, platform, digital identity and algorithmic decision-making, are cross-tabulated with domains of application, including the state, the market, citizens, intermediary organisations and crisis contexts. The grid functions both as an inventory instrument and as a means of identifying gaps: empty cells are explicitly assessed for their relevance.

**Reduction and delimitation.** The longlist generated in this manner is subsequently reduced on the basis of analytical utility. Episodes are combined when they form a single technological cluster with a comparable claim structure, and separated when their claim logic or type of transformation differs substantially. Policy codifications without independent regime effects are not included as primary episodes, but are assigned to a peripheral layer.

**Quadrant distribution.** The resulting corpus is provisionally positioned within the 2×2 matrix of predictive accuracy × realisation of transformation. This classification is analytically provisional: the definitive assessment takes place in Chapters 5–8. The purpose of the quadrant distribution is not to confirm success or failure, but to create conditions for comparison.

**Peripheral layer of the canon.** In addition to primary regimes, a peripheral layer is distinguished, consisting of corrective and normative interventions, such as privacy legislation, the institutionalisation of cybersecurity and oversight mechanisms. These are not placed within the 2×2 matrix, but function as second-order responses to earlier digital transformations.

The result is a transparent, systematic and verifiable corpus that provides the starting point for the case analyses in Chapters 5–8. Canon formation is not an endpoint here, but a methodological precondition for empirical analysis.

# 5. Quadrant Q1: Accurately Predicted and Realised

Chapter 5 analyses episodes in which explicitly articulated future-oriented claims correspond to a substantial degree with the societal transformation that was ultimately realised. This category is not a normative list of successes, but an analytical reference point: it reveals the conditions under which predictions prove structurally adequate.

The cases examined within this quadrant mainly concern core infrastructural and institutional regimes, including large-scale administrative automation at the Dutch Tax and Customs Administration and within the Municipal Personal Records Database, network development through SURFnet and broadband, and digital identification through DigiD. In these cases, the technological intervention was positioned as a solution to specifically articulated problems: the increasing scale of administrative processes, the need for reliable data exchange, or secure digital access to public services.

**Characteristics of the predictions.** The future-oriented claims in the cases within this quadrant are characterised by: (1) relatively clear problem definitions; (2) explicit causal chains; (3) the absence of inflated assertions about the scale of impact; and (4) attention to institutional boundary conditions. Predictive quality is based here not on visionary ambition, but on configurative precision: the claims remained close to the concrete domain of application and rarely exceeded the articulated level of scale.

**Characteristics of the transformations.** The realised transformations concern the durable reconfiguration of administrative infrastructures and digital access systems. These changes are deeply embedded institutionally and display lock-in effects. Their societal impact manifests itself primarily at regime or system level within specific domains, such as taxation, population registration and digital identification, without a broader morphological revolution having been either claimed or realised.

**Causal structures.** In these episodes, technology does not function as an autonomous driving force, but as a necessary component within a coherent institutional configuration. Causality is multicausal but relatively stable: organisational need, technological capacity and administrative implementation reinforce one another.

**Interpretation.** The cases in Chapter 5 show that structurally adequate predictions occur primarily when: (1) the domain of application is clearly delimited; (2) claims about scale are proportionate; (3) institutional dependencies are explicitly incorporated; and (4) the technology is not presented as a generic solution to diffuse societal problems. The cases within this quadrant demonstrate that predictability in the digital domain is possible, but mainly under conditions of institutional embedding and discipline in the articulation of scale. Quadrant Q1 therefore serves as the reference point for comparison with the other quadrants.

# 6. Quadrant Q2: Accurately Predicted, but Realisation Falls Short

Chapter 6 examines cases in which the original future-oriented claim was epistemically adequate—proportionate, scale-consistent and plausible—but in which the realised societal transformation fell short in depth, speed or institutional embedding. Quadrant Q2 therefore marks a crucial analytical category: an accurate diagnosis does not guarantee structural realisation.

**ERP technology and digital chain integration.** The promise of ERP systems concerned the integration of core processes, real-time information provision and the harmonisation of organisational silos. This claim was substantively coherent and formulated with an appropriate awareness of scale. Integration as a principle has indeed been realised: ERP systems now form the backbone of administrative and logistical processes in both private and public organisations. The realised benefits, however, proved less linear and less strategically transformative than the original articulation suggested. Implementations were costly, complex and organisationally conflictual. Institutional absorptive capacity proved more limited than anticipated. The claim was adequate, but its realisation remained fragmented and often incremental.

**Electronic Patient Record/LSP: digital healthcare information exchange.** An explicit system-level claim was formulated for the Electronic Patient Record: nationwide interoperability would improve the quality, safety and efficiency of healthcare. Epistemically, this claim was proportionate and technically plausible. Its societal realisation nevertheless remained limited. Political resistance, privacy concerns and institutional fragmentation within the healthcare sector prevented full implementation. Digital information exchange has been realised, but in a reduced and fragmented form through the LSP, the National Exchange Point. The transformation fell short of the scale and institutional depth of the original ambition.

**Platform work.** The rise of digital labour platforms was recognised relatively early in policy and academic contexts as a structural shift in the organisation of work. Claims concerning flexibilization, the transfer of risks and the redefinition of employment relations were substantively adequate and conditionally formulated. Institutional realisation, however, remains ambivalent. Although platformisation is clearly occurring, its legal and social embedding remains unstable and contested. Case law and policy responses follow the development, but institutional stabilisation remains incomplete. The prediction was substantively robust; institutionalisation is still in progress.

The three cases reveal a consistent pattern: (1) epistemic adequacy is present; (2) structural transformation remains dependent on institutional absorptive capacity; (3) multicausality and implementation complexity moderate realisation; and (4) political and legal configurations partly determine the outcome. For the central research question, this implies that predictive quality is necessary but not sufficient for societal reconfiguration. The cases in Q2 show that structural transformation is not a direct function of accurate prediction, but of configurative congruence between technology, governance and institutional structure.

# 7. Quadrant Q3: Inadequate Predictions and Durable Transformation

The cases analysed in Chapter 7 demonstrate that far-reaching societal reconfiguration may occur even when, during the introductory phase, no proportionate or scale-aware predictions were articulated. The dominant claims surrounding commercial internet access, general-purpose AI and the historical mainframe transformation were predominantly instrumental, fragmented or cautious in policy terms. Explicit system-level hypotheses concerning shifts in coordination or legitimation structures remained limited or were entirely absent.

In practice, however, a structural reorganisation of information and decision-making architectures occurred. Network logic became constitutive of market and administrative processes; the use of mainframes led to centralisation and stabilised administrative regimes; and generative AI is reshaping knowledge practices and assessment structures within already digitalised environments. These transformations did not manifest themselves as abrupt ruptures, but as cumulative shifts across several layers simultaneously—technical, organisational, legal and normative. It was precisely this layered character that complicated early system-level articulation.

For the central research question, this implies that predictability in the ICT domain depends not solely on technological foresight, but on the capacity to recognise the configurative reordering of existing structures as a coherent pattern at an early stage. When change presents itself as optimisation or capacity expansion within familiar frameworks, the constitutive trajectories of regime change remain analytically underexposed. The Q3 cases thus show that under-articulation is not an incidental failure of communication, but a structural characteristic of layered and cumulative transformations.

# 8. Quadrant Q4: Inadequate Predictions and the Absence of Durable Embedding

Chapter 8 examines cases in which the relationship between expectation and realisation proves fundamentally problematic. In these cases, the articulated future-oriented claim was epistemically underspecified or overstated in terms of scale and robustness, while the realised societal outcome likewise failed to produce durable structural embedding consistent with the original ambition. Quadrant Q4 therefore represents not a simple failure, but a systematic discrepancy between claim formation and institutional development.

The Dutch Childcare Benefits Scandal illustrates that technical plausibility does not amount to structural adequacy in claims concerning societal impact. The assumption that large-scale data linkage and risk selection would produce more objective and efficient enforcement proved insufficiently sensitive to legal boundary conditions, error margins and the effects of scaling up. Digital classification functioned constitutively within the logic of implementation, but the institutional configuration in which it was applied generated effects that had not been articulated proportionately. The discrepancy between the claimed rationalisation and the disruption actually produced points to the underspecification of conditions and an overestimation of scale robustness.

In the CoronaMelder case, the procedural functioning of the technology was largely consistent with its design, but the claim concerning societal impact proved dependent on behavioural and institutional boundary conditions that had been made explicit only to a limited extent. Its assumed substantial contribution to reducing transmission depended on adoption rates, compliance and organisational

embedding. The contribution actually realised remained supplementary and temporary, without producing durable effects in infectious-disease control. The inadequacy lay not in the technical functionality itself, but in the insufficient specification of the configuration within which effectiveness could plausibly be achieved.

The comparison between Viditel in the Netherlands and Minitel in France shows that comparable classes of technology may follow divergent institutional trajectories. In France, a coherent configuration of distribution, standardisation and tariff mechanisms led to durable embedding, whereas the ecosystem conditions in the Netherlands remained fragile. In the Dutch context, the future-oriented claim concerning interactive digital services was not supported by a congruent institutional infrastructure. The inadequacy therefore concerned not the feasibility of the system, but the underestimation of the scale and adoption conditions required.

The case of digital electoral infrastructure confirms that technical modernisation becomes structural only when it is congruent with the existing institutional architecture. In the Netherlands, electronic voting remained procedural and temporary; in Estonia, internet voting was embedded in an already established digital state infrastructure. This divergence demonstrates that claims concerning the impact of digital infrastructure are structurally adequate only when they explicitly incorporate the institutional configuration in which implementation takes place.

Taken together, these episodes show that inadequate predictions in the digital domain rarely result from obvious technical error. More often, they involve an overestimation of scale robustness, the underspecification of institutional boundary conditions and the insufficient articulation of multicausal configurations. The structural lesson of Q4 is therefore that an accurate assessment of technological possibilities is insufficient: predictive quality requires precision in identifying the conditions under which technology may become constitutive of durable reconfiguration. Quadrant Q4 shows that, when such configurative precision is absent, both claim and realisation lack structural coherence.

# 9. Cross-Case Analysis and Patterns of Predictive Quality

In Chapter 9, the perspective shifts from individual episodes to systematic comparison across the four quadrants. Whereas the preceding chapters analysed the relationship between expectation and transformation for each case, this chapter examines which characteristics of predictions and predictors are repeatedly associated with structural adequacy. The analysis does not seek to reclassify the episodes, but to identify recurring patterns in the articulation of claims, the outcomes of transformation processes and causal structures.

Within the epistemic dimension, structurally adequate predictions are characterised by the explicit elaboration of causal mechanisms, scale-aware positioning and conditional precision. Where system changes were articulated proportionately and boundary conditions were specified analytically, the relationship between expectation and realisation proved more robust. By contrast, the underspecification of conditions and the overestimation of scale robustness are associated with implementation outcomes that systematically fall short of the claims. Predictive power lies not in visionary rhetoric, but in configurative precision.

Analysis of implementation outcomes shows that the depth and durability of transformation vary according to the extent to which technology becomes constitutive of institutional architecture. Infrastructural regimes characterised by broad diffusion and strong lock-in effects prove relatively stable, whereas interventions that remain dependent on voluntary adoption or limited organisational

absorptive capacity rarely produce structural reconfiguration. The analysis confirms that regime formation is determined primarily not by intention, but by institutional embedding and cumulative dependence.

From a causal perspective, monocausal attribution is rarely tenable. Almost all episodes involve configurative causality, in which technological capabilities, legal frameworks, organisational capacity and political priorities mutually reinforce or constrain one another. Structurally adequate predictions are distinguished by their explicit recognition of this multicausality and by their refusal to reduce it to a linear technology–effect model. Where technology is positioned as an autonomous driver, the risk of derailment increases.

In addition to these patterns within primary regimes, the chapter considers second-order effects and peripheral corrective mechanisms. Digital infrastructures systematically generate normative, legal and security-related responses. Privacy legislation, cybersecurity architectures and oversight mechanisms emerge not as incidental additions, but as reflexive stabilisation mechanisms within advancing digitalisation. This serendipity is not random, but structurally probable, although difficult to predict.

For the central research question, the cross-case analysis implies that the predictability of IT-based societal transformation depends not primarily on technological foresight, but on insight into causal structures. Structurally adequate predictions combine scale consistency, the explicit specification of conditions and recognition of layered causality. At the same time, complete predictability remains constrained by emergent second-order effects and institutional dynamics. The ICT domain is therefore not fundamentally unpredictable, but it is systematically vulnerable to oversimplification.

# 10. Conclusions, Discussion and Recommendations

Chapter 10 brings the research questions together in a systematic set of answers and positions the findings within their methodological and theoretical limits. In answering RQ0, canon formation is understood not as an inventory, but as an epistemic filter that disciplines the field of assessment by restricting it to cases in which both explicit future-oriented claims and observable societal effects can be reconstructed analytically. The 2×2 quadrant does not function as an ex post classification, but as a predefined instrument of variation that prevents success bias and enables analysis independently of quadrant membership. The answer to the first research question shows that the structural adequacy of future-oriented claims is associated primarily not with temporal precision or rhetorical persuasiveness, but with proportionate articulation of scale, the explicit formulation of causal assumptions and the identification of institutional boundary conditions. Claims that situate their impact within recognisable institutional structures and make dependencies explicit display greater structural congruence with subsequent developments than claims that postulate system-level reconfiguration without further specification.

The answers to the second and third research questions establish that realised digital transformation occurs predominantly at regime and domain level, and that technology rarely functions as an autonomous or sufficient cause. Attribution therefore requires a configurative reconstruction of mechanisms and complementarities within open societal systems. The analysis of variation across the quadrants confirms that structural adequacy does not coincide with successful outcomes: realisation may fail despite adequate articulation, while transformation may occur without an accurate prior prediction. The answer to the fourth research question identifies recurring characteristics of robust prediction: discipline in the articulation of scale, explicit specification of causality and boundary conditions, avoidance of assumptions of linear diffusion, and restraint in relation to timing. The central

research question is therefore answered by the proposition that the structural adequacy of technological future-oriented claims has historically depended not on successful outcomes or temporal accuracy, but on the articulation of scale and causal assumptions, with boundary conditions being explicitly identified and remaining proportionate to the intended level of impact.

The subsequent discussion in §10.2 examines whether these conclusions are methodologically tenable. The execution of the study is evaluated against the principles formulated in Chapter 3: a historically grounded analysis based on abductive reasoning and directed towards configurations of causal factors.

The discussion explicitly addresses hindsight bias, retrospective coherence construction, selection bias and confirmation bias. These vulnerabilities are acknowledged as inherent in historical research, but are mitigated through transparent selection, explicit analytical categories and a strict separation between claim-oriented, ontological and causal dimensions. The validity of the findings therefore rests on conceptual clarity and the consistent application of predetermined criteria, rather than on experimental control. In §10.3, the identified limitations are translated into targeted recommendations for further research, including international comparison, quantitative operationalisation of predictive quality, actor-oriented approaches and prospective application of the assessment framework. These extensions do not imply revision of the core findings, but are intended to deepen and further test the analytical instruments developed in this study.

# 5. Management Summary of the dissertation 'Columba Loquens: Pattern Recognition, Navigational Robustness and the Limits of Explanation in Avian Intelligence'

## Chapter 1 (Introduction: Reliable Performance, Inadequate Explanations and the Need for Repositioning)

Chapter 1 positions the central problem of this dissertation: the tension between reliable behavioural performance and inadequate explanations of that performance. Within the natural sciences, systems may function consistently while the underlying explanations are only partially and satisfactorily articulated. In this dissertation, 'explanation' is understood not as a complete mechanistic reconstruction, but as a coherent set of assumptions that renders behaviour plausible and reproducible. The specific research domain concerns homing pigeons (Columba livia) as a biological model. These organisms display stable performance in pattern recognition and homing, including under variable or incomplete information conditions. This robustness makes the domain epistemologically relevant: performance remains consistent while multiple explanatory models are simultaneously compatible with the observed outcomes. Explanatory pluralism is therefore treated not as a temporary lacuna, but as a potentially structural feature of complex behavioural domains.

The chapter introduces four interrelated research questions that jointly determine the analytical trajectory of the dissertation: (1) classification and pattern recognition under variation; (2) decision-making under uncertainty; (3) homing as a field-based stress test for explanatory models; and (4) the identification of minimal explanatory assumptions. This structure makes clear that the aim is not to defend a single dominant theory, but to investigate which assumptions prove necessary to understand stable performance. Chapter 1 therefore performs a strictly positioning function. It contains neither a literature synthesis nor a methodological elaboration, but formulates the problem in a manner that justifies the subsequent empirical and methodological steps. The point of departure is clear: the core problem is not the absence of data, but the assessability of explanations when multiple models can accommodate the same behaviour.

## Chapter 2 – Literature Review

Chapter 2 surveys the existing literature on classification behaviour and homing navigation in pigeons in order to specify the current state of knowledge and delimit the empirical scope of the research. The chapter does not develop a new theory, but reconstructs provisional answers from the field and identifies where underdetermination occurs.

The literature on classification shows that pigeons can maintain stable performance under systematic variation of stimulus material. Generalisation to new or modified exemplars is well documented empirically. At the same time, these performances do not compel adoption of a single necessary internal explanatory framework. Multiple descriptions of structural sensitivity to stimulus regularities remain compatible with the same behavioural data. Performance sustainability constrains assumptions, but does not eliminate them. Research on decision-making under uncertainty generally shows gradual shifts in performance when information quality varies. All-or-nothing transitions are rare within ecologically plausible ranges. This indicates the simultaneous contribution of multiple information sources, without

empirically enforcing the exclusive dominance of any one source. Here too, model selection remains limited for as long as behaviour remains functionally stable. Within the homing literature, magnetic, olfactory, solar-based and visual lines of explanation have been developed, each supported by specific manipulation paradigms. Review studies nevertheless show that comparable behavioural outcomes often remain compatible with multiple models. Manipulations reveal sensitivity, but rarely produce reduction to a single necessary mechanism.

Taken together, the literature reveals a consistent pattern: robust performance sustainability under variation, combined with persistent explanatory pluralism. This defines the experimental scope of the dissertation. Chapter 2 positions the research by specifying what is known, where underdetermination occurs, and which forms of systematic variation are suitable for placing assumptions under pressure without presupposing model reduction.

# Chapter 3 – Methodology

Chapter 3 articulates the methodological framework within which the experimental space identified in Chapter 2 is actually exploited. Whereas Chapter 2 showed that multiple explanatory models of classification and homing behaviour can remain simultaneously empirically compatible with stable behavioural outcomes, Chapter 3 operationalises how this compatibility can be addressed systematically and under controlled conditions. The methodology is therefore not a neutral account of implementation, but an explicit translation of an epistemological problem into an empirical design. Central to this design is an assumption-driven, empirical-comparative strategy. Rather than testing a single explanatory model against alternatives, the core assumptions embedded in established models — such as the necessity of specific information sources, the dominance of particular cues or the hierarchical ordering of mechanisms — are made explicit and subjected to controlled variation. The purpose is not to identify a 'winning' model, but to investigate the conditions under which performance sustainability remains intact and which assumptions prove descriptively necessary to characterise that maintenance.

The methodological design is structured around four research questions with distinct statuses. Research Questions 1 to 3 are empirical and generate the data on which the dissertation rests: classification behaviour under systematic stimulus variation (RQ1), homing behaviour under separate variation of navigational cues (RQ2), and homing under combined variation of multiple information sources (RQ3). Research Question 4 has an analytical-theoretical function and organises the empirical findings at the level of assumption structures. This differentiation ensures that the description of empirical patterns and the assessment of models are not prematurely conflated.

Data collection is organised in such a way that behavioural outcomes are recorded under reproducible and controlled conditions. Classification studies employ repeated measurements using systematically varied stimulus sets, while homing research consists of field experiments involving controlled variation of environmental information. A defining feature of the design is its use of gradual, proportional variation. The research seeks to place assumptions under pressure without immediately destabilising the system. Variation therefore functions as an analytical instrument rather than merely as a disturbance. Data analysis is primarily descriptive and comparative. Data are summarised using proportions, means and measures of dispersion, while statistical techniques support comparisons between patterns. Crucially, statistical outcomes are not used as decisive criteria for model selection. The analytical step associated with Research Question 4 focuses on the relationship between empirical patterns and the assumptions presupposed by explanatory models: which assumptions prove unnecessary, which remain compatible, and which are actually addressed within the range investigated?

Validity and reliability are procedurally safeguarded. Internal validity is strengthened through consistent measurement procedures and explicit reference conditions; reliability rests on repeated measurements and transparent recording protocols. Particular attention is also given to interpretative validity: empirical

findings and theoretical evaluation are kept strictly separate, ensuring that conclusions remain proportionate to the design. The methodology is also explicitly embedded within ethical constraints. Variations are introduced incrementally and within ecologically plausible limits; extreme or irreversible manipulations are excluded. Ethics therefore functions not as an external assessment imposed afterwards, but as a constitutive condition of the design itself.

Chapter 3 thus forms the methodological link between the literature review in Chapter 2 and the empirical investigations in Chapters 4 to 6. It establishes that underdetermination is not circumvented, but must be investigated systematically.

## Chapter 4 – Classification under Controlled Stimulus Variation

Chapter 4 constitutes the first empirical elaboration of the methodological framework described in Chapter 3. Whereas Chapter 2 surveyed the experimental scope surrounding classification ability and Chapter 3 explained how assumptions could be made explicitly addressable, this chapter examines concretely how the classification behaviour of homing pigeons responds to systematic variation in stimulus exemplars. The central question is whether stable classification performance depends on the exact repetition of previously presented stimuli, or whether it is maintained when stimulus configurations change. The experimental design therefore introduces variation in the form, structure and context of visual stimuli within predefined task formats. In addition to training stimuli, new, non-identical exemplars are introduced to investigate the extent to which response distributions generalise.

The results show that classification performance does not collapse abruptly under these variations. The introduction of new exemplars produces measurable shifts in response frequencies, but not a complete disorganisation of behaviour. Performance sustainability is therefore not tied to exact stimulus identity. At the same time, the data do not compel any specific interpretation of the underlying processing mechanism. The empirical patterns are compatible with different descriptions of sensitivity to structural regularities, without any single form of representation being identifiable as strictly necessary. Importantly, the chapter does not claim to demonstrate the existence of classification ability — this is already beyond dispute in the literature — but investigates which assumptions are necessary to describe performance sustainability. The findings constrain strong claims concerning exclusive dependence on specific stimulus presentations, but do not reduce the explanatory space to a single compelling framework. Here too, underdetermination proves to be an empirically demonstrable phenomenon within the range investigated.

Chapter 4 thus demonstrates that stable behavioural outcomes can occur under variation without the underlying assumptions being determined unambiguously.

## Chapter 5 – Homing under Variation of Individual Navigational Cues

Chapter 5 constitutes the second empirical pillar of the dissertation and addresses Research Question 2: how does homing behaviour change when individual navigational information sources are systematically varied? Whereas Chapter 4 focused on classification under stimulus variation, the focus now shifts to field-based navigational behaviour under controlled manipulation of sensory information. The point of departure is that existing explanatory models often implicitly assume that specific information sources — magnetic, olfactory, visual or solar-based — are necessary or dominant for successful return. The research design operationalises these assumptions not by isolating a single mechanism, but by incrementally reducing or modifying individual cues within ecologically plausible

ranges. Homing performance is measured in terms of return rate, initial orientation deviation and return time.

The results show a consistent pattern of gradual sensitivity. Reduction or manipulation of an individual cue produces measurable shifts in performance indicators, but not a systematic collapse of goal-directed return behaviour. Within the range investigated, no single cue proves strictly necessary in an exclusive sense. At the same time, no cue is without influence: changes in availability correspond to proportional shifts in orientation and flight duration. The findings imply that assumptions concerning strict necessity or hierarchical dominance are not descriptively indispensable within this range of variation. Homing behaviour remains compatible with multiple configurations of information use. Explanatory pluralism is therefore not eliminated but constrained: strong exclusivity claims lose their force without alternative mechanisms being eliminated.

The chapter thus provides empirical evidence that goal-directed navigational behaviour can remain robust under partial uncertainty, while the underlying structure of assumptions is not reduced to a single mechanism.

## Chapter 6 – Performance sustainability under Combined Variation

Chapter 6 deepens the analysis of homing behaviour by moving from the manipulation of individual cues to cumulative variation. Whereas Chapter 5 examines sensitivity to individual information sources, this chapter addresses Research Question 3: how does the navigational system behave when multiple navigational cues are simultaneously restricted? The central analytical distinction is between analysis of individual elements and system behaviour under system pressure. Rather than isolating a single cue, multiple dimensions of variation are combined in accordance with predefined configurations. The primary purpose is therefore not to test which mechanism is operative, but to determine whether performance under cumulative uncertainty continues to shift proportionally or instead displays a non-linear transition.

The empirical findings show that combined reduction produces further performance degradation: the dispersion of orientation increases and return times become longer. Crucially, however, no abrupt collapse of homing behaviour occurs within the parameter ranges investigated. Even under increased system pressure, return performance remains above chance level and recognisably goal-directed. This outcome has two implications. First, it supports the interpretation of robustness: the navigational system displays resilience under cumulative uncertainty. Second, explanatory pluralism remains intact. Even under combined pressure, no single explanatory model can be enforced exclusively. Assumptions concerning threshold behaviour or necessary system collapse receive no empirical confirmation within this design, but neither are they excluded in principle beyond the range investigated.

This chapter therefore demonstrates that performance sustainability and explanatory pluralism occur together not only under isolated variation, but also under cumulative pressure. The system degrades proportionally, without compelling reduction to a single mechanism.

## Chapter 7 – Answering Research Question 4

This chapter shifts from empirical manipulation to analytical evaluation. Whereas Chapters 4 to 6 documented behavioural outcomes under controlled variation, this chapter addresses Research Question 4: which assumptions about internal processing prove necessary to understand the observed performance, and which can be relinquished without loss of explanatory coherence? The analysis begins with the combined empirical patterns. In both classification (Research Question 1) and homing under individual and combined variation (Research Questions 2 and 3), performance shifts proportionally without abrupt discontinuity within the range investigated. No individual information source proves strictly necessary, and cumulative reduction does not lead to system collapse.

These findings place strong assumptions under pressure. Assumptions concerning exclusive necessity or hierarchical dominance lose their descriptive force when behaviour is maintained despite the reduction of individual or multiple cues. At the same time, the models concerned are not eliminated. The data neither exclude specific mechanisms nor render them unavoidable. The central outcome of this chapter is therefore a differentiation between the statuses of assumptions. Some assumptions prove unnecessary within the chosen design; others remain compatible but are not exclusively confirmed; still others remain empirically undecidable. Explanatory pluralism is therefore not reduced, but explicitly ordered.

The chapter concludes that model assessment in this domain shifts from performance comparison to assumption analysis. Assessability requires assumptions to be made explicit and their descriptive necessity to be tested. Within the ranges of variation investigated, pluralism persists not as a theoretical weakness, but as an empirically bounded finding.

## Chapter 8 – Conclusion, Discussion and Further Research

Chapter 8 is primarily concerned with synthesis and reflection. The four research questions are answered individually and subsequently brought together in answering the central research question. With regard to classification behaviour (Research Question 1), performance remains stable under systematic variation of stimulus exemplars. Generalisation occurs without the data compelling reduction to a single necessary representational framework. With regard to homing under individual variation (Research Question 2), return rates and return times shift gradually, without any single cue being identifiable as strictly necessary or exclusively dominant. Under cumulative variation (Research Question 3), performance degradation becomes apparent, but no abrupt collapse occurs within the parameter ranges investigated. Robustness proves bounded, but persistent. On this basis, Research Question 4 is answered: explanatory models remain difficult to eliminate in relation to one another for as long as they remain compatible with observed performance sustainability. Making assumptions explicit narrows the scope for strong exclusivity claims, but does not lead to reduction to a single dominant model. Explanatory pluralism thus emerges as an empirically bounded feature of the relationship between behavioural stability and explanatory structure.

The discussion qualifies this outcome. Underdetermination is not presented as a trivial property of complex systems, but as a finding within specific ranges of variation and particular behavioural measures. Internal validity is strong within the chosen design, but the scope of the findings is constrained by ethical proportionality, measurement resolution and population variation. The dissertation concludes that reliable behavioural performance is possible without enforceable explanatory reduction. Stability and underdetermination prove not to be opposites, but simultaneously occurring features of the same system. The epistemological implication is formulated proportionately: within this domain and under these conditions, performance sustainability does not coincide with explanatory reduction.