Department of Computing Science, 2026

# Contextual Language Processing: From Formal Transducers to Contextual Advertising

Jingwen Cai


UMEÅ UNIVERSITY

ISBN: 978-91-8070-908-8 (print)
ISBN: 978-91-8070-909-5 (pdf)
ISSN: 0348-0542
Report / UMINF no. 26.01

Cover: "Contextual Communication Between Intelligences", designed by Jingwen, with support from Doubao, ChatGPT, and Photopea.

# Abstract

Contextual language processing concerns how information from the surrounding media environment, such as accompanying texts or interaction signals, is represented and used in automated language processing systems. Incorporating contextual information enables models to derive interpretations that go beyond isolated textual analysis and capture aspects of meaning grounded in real-world situations. Despite substantial research in language modeling, effectively processing contexts remains an appealing challenge, particularly in balancing the expressive power, interpretability, and adaptivity of the model, and its alignment with human understanding.

This thesis investigates contextual language processing from both methodological and applied perspectives, focusing on two related domains: natural language processing and contextual advertising. On the methodological side, the thesis formalizes the grammatical inference of finite-state transducers as a sequential decision-making problem. By integrating structured transducer models with reinforcement learning, it demonstrates how contextual languages can be learned dynamically while retaining interpretability. These learning-based approaches offer a systematic way to adapt language transformations to shifting contexts. Such adaptivity can, for instance, lead to more effective decisions in automated advertising auctions. On the applied side, the thesis examines contextual communication in real-world, user-facing systems through empirical studies in contextual advertising. These studies reveal a gap between the computational level of contextual relevance that is often optimized by automatic metrics, and human-perceived relevance as experienced by end users. The findings show that increased algorithmic precision or model complexity does not necessarily translate into improved user experiences, and that factors such as transparency and trust play a central role in contextual effectiveness.

Through theoretical, empirical, and conceptual analyses, this thesis demonstrates that effective contextual language processing necessitates bridging multidisciplinary perspectives, including algorithmic and perceptual ones. It frames contextual language processing as an interpretable, adaptive, and user-centered interaction process. The insights discussed in this work offer implications not only for algorithmic optimization, but also for modern online advertising and other user-facing applications where contextual understanding, practical constraints, and privacy considerations are critical.

# Preface

This thesis investigates contextual language processing across two domains: formal transducers and contextual advertising. It includes the following papers:

Paper I Johanna Björklund, **Jingwen Cai** and Anna Jonsson. Reinforcement Learning of Finite-State String Transductions. *Journal of Automata, Languages and Combinatorics*, 29, 2–4, 109–136, 2024.

Paper II **Jingwen Cai**, Johanna Björklund and Frank Drewes. Reinforcement Learning of Probabilistic Finite-State Transducers. Manuscript, 2025.

Paper III **Jingwen Cai**, Sara Leckner, and Johanna Björklund. From precision to perception: Human-in-the-loop evaluation of keyword extraction for internet-scale contextual advertising. *Information Systems*, 138, 102665, 2026.

Paper IV **Jingwen Cai**, Bart Piet Knijnenburg, Johanna Björklund and Sara Leckner. Beyond Precision: Understanding the Impact of Algorithmic Accuracy and Transparency on User Perceptions in Keyword-Driven Contextual Advertising. Manuscript; submitted to the *International Conference on Human-Computer Interaction*, 2026.

Paper V Adam Åbonde Garke, Richard Wahlund, Sara Leckner, Emil Häglund, **Jingwen Cai**, Igor Ryazanov and Johanna Björklund. Programmatic Advertising in the Age of AI: A Conceptual Overview and Strategic Recommendations. Manuscript, submitted to the *International Journal of Advertising*, 2025.

Although three papers included in this thesis are manuscripts at the time of thesis submission, they present part of the core contributions and the most important findings of the thesis. Taken all the included papers together, they form a coherent body of work that directly addresses the intended research aims. In addition to these papers, the following papers were published within the studies but are not included, as some represent preliminary versions, while others address complementary topics that fall outside the central scope of this

thesis. Although not included, these works contribute to the development of the studies presented here and help shape the overall research direction.

Paper VI Hannah Devinney, Anton Eklund, Igor Ryazanov and **Jingwen Cai**. Developing a multilingual corpus of wikipedia biographies. *In Proceedings of the 14th international conference on Recent Advances in Natural Language Processing*, pages 285–294, 2023.

Paper VII **Jingwen Cai** and Johanna Björklund. Exploring Contextual Advertising in Real-Time Bidding: Evaluating RL Algorithms under a Constrained Environment. *Proceedings of the Workshop on Strategic and Utility-aware REcommendations at the 18th ACM Conference on Recommender Systems*, 2024.

Paper VIII **Jingwen Cai** and Johanna Björklund. Reinforcement Learning in Online Advertising: Challenges, Prospects, and Trust. *Proceedings of the Workshop on Reliable and Trustworthy Artificial Intelligence at the 16th Asian Conference on Machine Learning*, in press, 2025.

Paper IX **Jingwen Cai**, Sara Leckner, and Johanna Björklund. From Precision to Perception: User Surveys in the Evaluation of Keyword Extraction Algorithms. *Proceedings of the Workshop on Human-In-the-Loop Data Analytics at the International Conference on Management of Data*, Article 2, 1–9, 2025.

This doctoral study was funded by the Marianne and Marcus Wallenberg (MMW) Foundation under grant number 2020.0095. The thesis was written as part of the multidisciplinary research project on "AI-driven contextual advertising" financed by MMW, involving researchers from computer science, media studies, and marketing.

# Acknowledgements

I am not naturally expressive, and I often find it difficult to articulate my emotions. Yet I carry deep gratitude for the people I have met along this journey and for those who have supported me along the way. I know that I would not have come this far without them. Although writing these down is easier for me than expressing such feelings aloud, my gratitude goes far beyond what words can fully convey.

First and foremost, I would like to express my sincere gratitude to my supervisors, mentors, and advisors, whose guidance has shaped both my academic development and my growth as a researcher. My main supervisor, Johanna Björklund, has supported me with remarkable generosity, not only in research but also with generous care and understanding. Her wisdom, energy, and confidence have been a true role model to me. Through working with her, I have learned not only how to conduct research, but also how to reflect critically and independently. The freedom and trust she placed in me allowed me to explore my research interests openly, while her insightful guidance consistently helped me think more deeply and stay on the right path. In many ways and for part of the motivation I started my doctoral studies here, she embodies the aspiring vision I carried at first, and I am profoundly grateful for her influence. I am also deeply thankful to Sara Leckner, my co-supervisor, whose kindness and encouragement have meant a great deal to me. Her insightful comments and precise suggestions repeatedly helped sharpen my research. Beyond her role as a supervisor, she has been a supportive mentor and a valued friend. I feel very fortunate to have attended the RecSys'24 tutorial, where I met Bart Piet Knijnenburg and had the opportunity to collaborate with him. His expertise in user studies and statistical analysis inspired and challenged me intellectually and deepened my interest in this field. I greatly enjoyed learning from him and being enlightened beyond my comfort zone. I would also like to thank Frank Drewes and Anna Jonsson, who deepened my understanding of the beauty of formal languages. The discussions with them have strongly influenced my methodological thinking and helped cultivate a rigorous approach to research and problem formulation.

I am immensely grateful to my other research family members, who welcomed me with warmth and generosity and made this academic journey truly meaningful. Rahil and Willeke, the members of our clumsy club, have always

cared for me and shared countless joyful moments. Igor and Lena have been genuine friends, always willing to offer help and support whenever needed. Emil and Anton, wonderful companions with whom I shared an office for most of my daily work. Štěpán, Martin, and Henrik (of course Johanna as well), I am especially thankful for your kindness and help when Noodles needed surgery. Adam and H, thank you for welcoming me so warmly during my first days here and for your continued kindness and friendship. Arezoo and Sal, your kindness has meant a lot to me. Thinking of leaving you all is perhaps the most painful aspect at this point. The four and a half years we spent together have been among the most precious experiences of my life here as an international student. You made a new country feel like home.

I would also like to thank my Chinese friends here, who created spaces where we could share joy, relax, and momentarily step away from research pressures. You were my comfort whenever homesickness set in. Though people come and go and we each face our own academic and personal challenges, the parties we shared, the board games we played, and the dinners we cooked together helped ease that distance from home.

To my beloved parents, 妈妈爸爸, you are the foundation and confidence behind everything I do. You have basically devoted yourselves selflessly to me. and because of you, I am now able to stand on my own. It is now my turn to take care of you and to support you as you truly enjoy your lives. 姥姥, my greatest concern is not being able to stay with you as often as I wish. At the same time, I am also grateful that you are happy and in good health. Please know that you are always in my thoughts and close to my heart. 姥爷, I know you would be proud of me. To all my other family members, thank you for your unwavering love and support. Your care is what sustains me.

Finally, thank you, Ding and Lena, for standing by me during my darkest days. I could not have made it through without you. And to my dearest Noodles, who loves me unconditionally with his whole little heart, a new adventure lies ahead, and I know I will never walk it alone.

# Contents

# Chapter 1

# Introduction

## 1.1 Motivation

Contextual information plays a central role in how language is interpreted and processed in modern Artificial Intelligence (AI) systems. From Natural Language Processing (NLP) tasks, such as machine translation and information retrieval, to user-facing applications including contextual advertising and online recommendation, the performance of these systems critically depends on how contextual information is identified, represented, and used. Despite substantial progress in both theoretical foundations and algorithmic development, effectively processing contexts remains a persistent and challenging problem.

Early approaches to contextual language processing were largely based on structured, rule-based models, which offer formal clarity and interpretability but are limited in their ability to capture rich or evolving contextual dependencies. More recently, neural and deep learning-based models have significantly expanded the expressive power of contextual representations, enabling improved performance on a variety of benchmark tasks. However, these advances have also raised concerns regarding interpretability, controllability, and the transparency of their outcomes and behaviors. In particular, the reliance on black-box neural architectures complicates the understanding of how contextual information influences system outputs. These concerns become especially pronounced in real-world, user-centered applications. Phenomena such as hallucinated or unstable responses produced by Large Language Models (LLMs) have highlighted the risks associated with unconstrained, opaque language processing, undermined user trust and credibility, and have intensified calls for more explainable and controllable solutions.

At the same time, the evaluation of contextual language processing models has largely relied on automatic metrics and predefined benchmarks that prioritize precision and efficiency. While such metrics provide standardized means of comparison, they do not necessarily reflect whether contextual decisions are meaningful, effective, or reliable from a user perspective. In domains

like contextual advertising, contextual understanding extends beyond identifying increasingly precise contextual signals about the content or the user, but also involves how contextual relevance is perceived and reacted by users. This combined perspective is under-explored due to its multidisciplinary nature and the rapid evolution of its techniques, and this thesis contributes to addressing this issue. Accordingly, contextual language processing cannot be reduced to purely technical optimization, and algorithmic performance and user-perceived contextual effectiveness cannot be equated. Improvements in computational metrics do not automatically guarantee user preference or engagement.

These limitations point to a broader gap in contextual language processing research. On the one hand, there remains a clear methodological gap between structured, interpretable models and adaptive learning-based approaches. On the other hand, there is a practical divide between algorithmic optimization and user perception of contextual understanding. Addressing these challenges requires viewing contextual language processing as an interaction process shaped by various aspects, including modeling choices, learning objectives, resource constraints, and user interpretation.

Motivated by these challenges, this thesis investigates contextual language processing from both methodological and applied perspectives. On the methodological side, it formulates the grammatical inference of formal language models, represented by string-to-string Finite-State Transducers (FSTs), as an adaptive sequential decision-making process learned by interpretable Reinforcement Learning (RL) methods. On the applied side, it examines contextual communication in real-world scenarios through user studies in a keyword-based contextual advertising setting, demonstrating a long-overlooked gap between computational relevance and user-perceived relevance across a range of models, from traditional approaches to LLMs. Together, this thesis provides conceptual and empirical insights into how contextual language processing systems can be designed to become adaptive, interpretable, and aligned with human understanding in practical applications.

## 1.2 Research Aims

The aims of this thesis are as follows:

**AIM 1.** Advance contextual language processing through interpretable, learning-based models and user-centered evaluation in AI-driven applications such as contextual advertising.

**AIM 2.** Strengthen the theoretical and methodological foundations of contextual language processing by designing adaptive learning algorithms.

**AIM 3.** Bridge algorithmic and perceptual perspectives in contextual language processing to better align computational performance with human understanding, considering constraints and implications for AI-driven advertising practices.

The papers summarized in Chapters 4 and 5 demonstrate how they contribute to these aims.

## 1.3 Thesis Outline

The thesis is organized into five sections, followed by five research papers. Section 2 provides the broader research background, and Section 3 formalizes the two central problems investigated in the papers and outlines the key methodologies used to address them. Section 4 summarizes the main contributions of each research paper included in this thesis. Finally, Section 5 discusses the findings and provides directions for future work.

# Chapter 2

# Background

Contextual language processing can be broadly viewed as leveraging more contexts, such as scenarios, semantic relationships, and background knowledge, when automatically processing language (e.g., texts) with algorithms [63, 102]. Incorporating such contextual information enables systems to derive meanings that are more appropriate and accurate than those obtained from processing words or texts in isolation. This thesis primarily focuses on two research fields closely related to contextual language processing. The first is natural language processing, where the main contribution lies in inferring finite-state string transducers and evaluating keyword extraction methods. The second is contextual advertising, a multidisciplinary field that relies heavily on marketing, psychology, and natural language processing techniques.

## 2.1 Natural Language Processing

### 2.1.1 Foundations of Language Processing

Natural language processing refers to learning, understanding, and producing human language using computational techniques [59, 87]. It began in the 1950s at the intersection of artificial intelligence and linguistics [18, 104, 128]. Over time, it has expanded into a substantially broader research area that now incorporates knowledge from the social sciences, psychology, cognitive science and several other related disciplines. Typical and common natural language processing tasks are for example, document clustering, information retrieval and speech recognition. The input and output processed in NLP tasks can be symbols, texts, sentences and tokens.

In the early stages of NLP research, simple approaches such as word-for-word machine translation were designed to "automate" language processing tasks. Researchers manually specified vocabularies and grammar rules in an attempt to make computers understand human language [101]. However, the complex, ambiguous, and context-dependent nature of human language re-

vealed the limitations of these early rule-based approaches [59]. To enable a more systematic theoretical analysis of linguistic structure, Noam Chomsky introduced a series of idealized simplifications and abstractions, known collectively as the *Chomsky hierarchy*, to investigate the computational basis of human language [24, 39, 66, 101]. This framework formalized the relationships between different classes of formal languages, providing foundational insights into the complexity and computability of them. It is a cornerstone of formal language theory and continues to influence modern NLP research.

Within this framework, *regular languages* are defined at the lowest level, representing a class of languages with limited expressive power, yet efficient and well-structured, that can be recognized by *finite-state automata*. This establishes a clear boundary on the set of languages that can be processed under strictly limited computational resources and provides a controllable and computable starting point for the development of practical language processing methods. They formalize languages as symbolic sequences and enable the modeling of structured, finite, and local contextual constraints through sequential processing. This motivates the study of sequence-based representations and transformations for capturing task-relevant contextual patterns.

### 2.1.2 Sequence-Based Processing and Transformation

Building on this perspective, a natural language sentence can be viewed as an ordered sequence of symbols or tokens. Its processing can thus be treated as a type of sequence-based rewriting, in which input symbols are examined sequentially and decisions regarding interpretation and transformation are made incrementally. Therefore, many NLP tasks can be formulated as mappings from an input sequence to an output sequence, including machine translation [103]. One of the original and influential formal frameworks for such sequence-to-sequence processing is *finite-state transducers*, which define explicit, state-based mechanisms for realizing rational mappings between the source (input) and the target (output) symbol sequences [21, 80, 133]. Due to their favorable computational properties, FSTs offer strong advantages in terms of time and space efficiency [21, 99], which has made them particularly attractive for practical language processing systems, even as more powerful models exist.

From the learning perspective, a significant line of research has focused on the problem of *grammatical inference*, which concerns learning the formal grammars or transformations from data [9, 43]. In the finite-state setting, this involves inferring automata or transducers from observed samples, such as paired input-output strings. The goal in this context is to induce a state-based sequence transformation model that is consistent with observed translation pairs, thereby enabling data-driven, automatic construction of sequence processing systems. Early approaches explored gradient-based and Expectation-Maximization optimizations, yet they face challenges related to computational complexity [10, 25, 34, 114].

A related line of work has examined the close connection between weighted

FSTs and Hidden Markov Models (HMMs), particularly in speech recognition. In this context, the weighted FSTs provide a compact and computationally efficient representation for HMMs (see e.g., [20, 33, 44, 73, 100, 111]). In FSTs, the transition between input and output sequences is captured implicitly through an unobserved latent alignment. When weights are introduced, weighted FSTs associate each transition with a value, enabling the transducer to compute a mapping from pairs of input-output strings to a domain of, for example, costs, rewards, or probabilities [100]. Correspondingly, HMMs provide a mathematical framework that represents these latent alignments with hidden states and computes sequence probabilities by summing over all possible alignment paths (or searching for the optimal path/output during decoding). In addition, under the Markov assumption, sequence transformation depends on the current state, which provides a conceptual basis for learning approaches that operate over state transitions.

Later, with the advent of neural networks, the HMM-based transducer framework was retained, while the transition probabilities were replaced by neural parameterizations, leading to modern neural transducer models, including Connectionist Temporal Classification, Recurrent Neural Networks Transducers, and Transformers Transducers (see, e.g., [48, 49, 50, 57, 67, 97, 105, 113, 143]). Despite differences in modeling assumptions and expressive power, these approaches continue to share a common foundation in sequence-based processing of language transformations.

### 2.1.3 Contextual Language Understanding

As sequence-based approaches were applied to increasingly complex tasks, attention gradually shifted toward how contextual information could be more effectively incorporated to improve language processing. This shift led to a broader conception of context as information accumulated dynamically during processing, rather than as uniform or fixed references. Importantly, this evolution did not imply that all available information should be treated as equally relevant context. Instead, it motivated a much broader perspective that considers how relevant contextual information is selected and represented, how contextual usage is optimized under constraints of computational resources and learning objectives, and what constitutes an effective method for conveying meaningful contextual information in practice.

> *Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.*

This operational definition was introduced by Dey [30] for context-aware applications within the domain of human-computer interaction. It has significantly influenced a wide range of research areas, motivating increased interest in understanding and analyzing context ever since. In NLP, contextual language

understanding is more specifically concerned with how contextual information influences the interpretation and use of languages [7]. Early NLP approaches primarily focused on local and surface-level contextual cues, often representing them with isolated and discrete words or phrases based on their occurrence statistics, such as n-gram and dictionary methods [75, 121, 122, 124]. These approaches model context in a highly localized and explicit manner, capturing limited dependencies based on word proximity or frequency. However, despite their simplicity, such methods are still widely used, particularly in resource-constrained or task-specific settings due to their robustness, interpretability, and computational efficiency [2, 112].

Subsequent developments introduced distributed representations that encode contextual information implicitly through vector spaces. These word embedding methods represent lexical items as vectors, enabling the modeling of semantic similarity, relational structure, and latent contextual patterns beyond surface-level contexts [4, 75, 98, 121]. Within this paradigm, context is no longer defined explicitly, but can be discovered from learned representations dependent on training data and task objectives. More recent approaches further extend this perspective by modeling context as a dynamic construct that guides language processing decisions. Specifically, modern neural architectures emphasize the selective incorporation of contexts (see, e.g., [29] and [132]), which has played a key role in the rise of transformer-based models. However, the opacity of these powerful models has drawn growing concerns [60, 86, 144], as unresolved issues of trust and control are driving demands for explainable and interpretable solutions.

As the capacity of modern language models, and the depth and range of contextual information they process continue to increase, research has also focused on optimizing these models towards task-dependent and domain-specific objectives [6, 88]. Large language models are adapted or fine-tuned to better support particular tasks and application settings. In addition, frameworks such as in-context learning and chain-of-thought prompting aim to align these model behaviors more closely with user expectations and interaction needs [32, 135]. As a result, contextual understanding has shifted from merely maximizing contextual coverage towards identifying and conveying contextual information that is meaningful and effective in practical scenarios. This shift is not confined to NLP alone, but is also evident across a range of other domains and applications, as will be illustrated in the next section on contextual advertising.

## 2.2 Contextual Advertising

### 2.2.1 From Behavioral Targeting to Contextual Targeting

Throughout the history of advertising, a central objective of advertisers has been to deliver advertisements to suitable audiences [65]. This objective has long been pursued through various forms of context matching, where the concept of "context" is not new. Such context-matching strategies reflect the

observation that consumers are never isolated from their surrounding context when encountering an ad [55, 123]. In the early stages of advertising, context was primarily inferred from content and placement, with advertisers relying on the environment and thematic relevance of the content being consumed to determine where ads should appear. With the rapid development of online media platforms and advertising technologies, targeting strategies have expanded significantly in both scale and efficiency, increasingly driven by AI-based systems that support automated and large-scale advertising decisions. The success of platforms such as Google AdWords has demonstrated the feasibility of automated advertising systems capable of reaching broad audiences with improved precision and operational simplicity [142]. In addition, the widespread adoption of third-party cookies further enabled advertisers to access user-level behavioral data, opening the gate to data-driven personalization at unprecedented scale. As a result, *behavioral targeting* emerged as a dominant paradigm, in which major online and social media platforms leverage user browsing histories, search activities, and interaction patterns to construct detailed user profiles and then infer individual interests [12, 139, 142].

However, at the same time, behavioral targeting has raised growing concerns regarding privacy, data protection, and ethical use of personal information [1, 131]. The collection and use of these behavioral data have been widely criticized for their potential to infringe upon user privacy and autonomy. These concerns have translated into increasing legal and regulatory scrutiny, with advertisers facing great challenges in complying with evolving privacy standards. High-profile cases involving the mishandling of user data by large technology companies have further intensified public and regulatory attention to these issues [106]. In response, data protection regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) have introduced stricter requirements governing the collection, processing, and use of personal data [17, 115]. Although these regulations originate from different jurisdictions, they collectively reflect a broader global trend toward enhanced user consent and control in online advertising. In addition, major technology platforms have reassessed the role of third-party tracking strategies. Notably, Google announced in early 2020 its plans to phase out third-party cookies [95]. Although these plans have undergone multiple revisions, delays, and were eventually not fully implemented as initially proposed [53, 95], the underlying implications remain clear: reliance on extensive third-party behavioral tracking is increasingly constrained, and future advertising ecosystems are expected to place greater emphasis on privacy-preserving and controllable alternatives.

Correspondingly, increased attention has been shifted toward *contextual targeting* as a complementary alternative to behavioral targeting. Although contextual targeting has never disappeared from advertising practice, recent regulatory, technical, and societal changes have brought it back to the center of the stage. Modern contextual targeting approaches extend beyond simple content matching to incorporate richer, situational information such as page se-

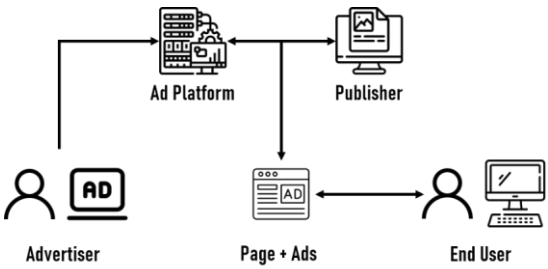



Figure 2.1: Illustration of a contextual advertising framework.

mantics, temporal factors and coarse-grained contextual cues, remaining largely non-identity-based [123]. Unlike behavioral targeting, which relies on long-term tracking and profile construction, contextual targeting emphasizes real-time, situational relevance derived from the current environment in which an ad is displayed [91]. This focus on limited, real-time, and observable context makes contextual targeting a more privacy-preserving, controllable, and transparent advertising strategy.

### 2.2.2 Principles of Contextual Advertising

Figure 2.1 illustrates a typical contextual advertising framework, which generally consists of four major components [1, 14, 110, 138]:

- *The subject* is an Internet user who visits a web page and may be exposed to an ad.
- *The publisher* owns or operates the web page on which the ads are displayed, aiming to provide a good user experience and maximize advertising revenue.
- *The advertiser* provides ads to promote their products or services, typically competing for ad placements through advertising campaigns.
- *The ad platform* serves as an intermediary between publishers and advertisers, collecting contextual information, operating ad campaigns and selecting suitable ads to display.

When a user opens a web page, an advertising opportunity is created, allowing advertisers to place their ads alongside the displayed web page content. Upon exposure, the user may choose to interact with the ad, for example, by clicking on it, which typically redirects the user to a landing page for more information. While this framework may be representative of modern programmatic advertising systems that often incorporate real-time bidding mechanisms [107, 117], contextual advertising is conceptually different. Its defining characteristic lies in how ads are matched based on contextual information, rather than

in the specific market mechanisms used for pricing or allocation. As such, it does not necessarily require real-time auctions, and the framework depicted in Figure 2.1 should be viewed as a general abstraction of contextual ad selection.

Prior studies have shown that higher ad-content relevance is associated with increased user engagement [70, 120, 127]. Therefore, in contextual advertising, advertisements are preferably selected to align with the content and situational context of the web page, to enhance user experience and increase the likelihood of meaningful engagement. Research within this field has largely focused on identifying which contextual information is most influential, how such information affects outcomes, and how it can be effectively applied in advertising strategies (see, e.g., [36, 51, 109]). Although this emphasis parallels developments in natural language processing discussed in the previous section, contextual advertising places greater focus on practical principles shaped by its application-driven and cross-disciplinary nature.

Advances in NLP have nonetheless played a significant role in enabling richer and more effective use of contextual signals in advertising systems. Contextual advertising methods have evolved from simple keyword-based matching approaches [68, 129], to topic and semantic modeling techniques that capture broader contextual meaning [14, 110], and more recently to enriched representation-based context modeling [137, 138]. Across these developments, the central objective has remained consistent, which is to optimize the use of contextual information to support advertising goals. These goals are typically concretized and assessed through a range of evaluation metrics, including measures of contextual relevance, click-through rate, and user engagement measures [46, 82, 116]. In modern advertising systems, these metrics are increasingly optimized and monitored by AI-driven models that automatically learn from large-scale interaction data [83]. Naturally, these metrics also serve as practical indicators of how effectively contextual information contributes to the ad performance. Therefore, despite differences in formulation, such metrics can be interpreted as proxies for contextual relevance among the ad, the web page, and the user.

However, system-level measures of relevance do not always align with how relevance is perceived by users [16, 19, 94]. In addition, the interpretability and transparency of contextual ad placements can substantially influence user responses and their attitudes towards the ad [76], particularly in light of ongoing concerns in terms of privacy and ethical advertising practices. These considerations motivate a shift beyond purely algorithmic notions of contextual matching toward more user-centered, effective, and trustworthy perspectives in contextual advertising strategies.

### 2.2.3 Relevance, Transparency, and User Perceptions

In the advertising literature, the bidirectional relevance between an advertisement and its surrounding content is commonly referred to *as ad-content congruence*, which describes the degree of semantic or contextual similarity between the two [36, 45, 109]. While some studies define congruence as a multidimen-

sional concept encompassing both relevancy and expectancy [58, 78], this thesis focuses primarily on the well-discussed relevance dimension, as it directly relates to contextual matching and language-based interpretation. A widely cited theoretical explanation for the effectiveness of ad-content relevance is the *priming effect* [118]. According to it, ads presented in semantically associated contexts can function as cognitive primes, activating associated concepts and facilitating memory retrieval to related topics [27, 77, 125, 141]. This increased processing fluency is often associated with more favorable cognitive and persuasive outcomes, such as clicks and purchases. However, empirical evidence suggests that the effects of ad-content relevance are not uniformly positive and may depend on various factors.

One of the challenges lies in the distinction between system-level relevance and user-perceived relevance. Although ad-content relevance can be formally defined, measured, and evaluated by advertising systems, its perceived effectiveness may vary substantially across users and situations. Prior studies have shown that factors such as user viewing goals, advertisement execution styles, and levels of involvement can significantly moderate the impact of relevance [36, 120, 125]. Under such conditions, the relationship between relevance and ad effectiveness is not necessarily linearly positive, and highly relevant ads may yield diminishing or even negative effects. Moreover, the semantic associations activated by the ad and the web page content can differ markedly across individuals [92, 140], suggesting that a uniform, system-defined measure of ad-content similarity may not reliably capture real user-level interpretations.

In addition, highly relevant ads may also raise privacy-related concerns and trigger avoidance behaviors [8, 70]. One theoretical explanation for this is provided by the *Persuasion Knowledge Model* (PKM), which posits that users interpret and respond to advertising attempts based on their understanding of persuasion tactics [42]. From this perspective, when ads appear excessively relevant, users may become more aware of the underlying persuasion intent and perceive a loss of freedom of choice, leading to defensive responses to the ad. Particularly, in modern online advertising environments, where AI-driven personalization and data usage are under ongoing public debate, high ad-content relevance may further amplify skepticism by signaling implicit personalization or surveillance [3, 23]. To address these issues, advertisers and researchers have increasingly explored transparency mechanisms that aim to explain why a particular ad is shown in a given context [3, 5, 76]. Such explanations are intended to align system decisions with user understanding and to mitigate negative perceptions. However, empirical findings suggest that transparency is not always beneficial. Excessive or poorly designed explanations may heighten perceptions of persuasive intent, increase privacy concerns, or backfire by drawing attention to ad practices that users find uncomfortable [3, 72]. The impact of transparency therefore depends critically on how explanations are designed and presented, including their level of detail, personalization and clarity [11, 35, 76]. Nonetheless, the research on how transparency impacts user perceptions and behaviors on the ad is still under-explored.

Overall, user perceptions of relevance, transparency, and trust play a central role in determining advertising outcomes, yet remain difficult to capture through automated metrics. This gap motivates a more user-centered analysis of how contextual relevance is perceived and responded to by users.

# Chapter 3

# Problem Formalization and Methodology

This chapter formalizes the main challenges discussed in this thesis: the design of adaptive mechanisms for learning finite-state string transductions, and the investigation of contextual relevance in real-world applications through contextual advertising. The methodological approaches chosen to address these problems are also outlined in this chapter.

## 3.1 Finite-State Transducers

A finite-state transducer is a formal model for mapping input strings to output strings through a finite set of states and transitions. It provides a structured mechanism for language modeling, where the current state encodes relevant contextual information that guides how input symbols are transformed. As a foundational device in NLP, FSTs have supported various tasks due to their interpretability and robustness, but their behavior is typically constrained by statically specified rules.

Papers I and II focus on the grammatical inference of string-to-string transducers in the non-probabilistic and probabilistic settings, respectively. Grammatical inference shifts the source of transduction rules from predefined specifications to data-driven learning, thereby enabling the learned transducer behavior to be adapted as language and contextual patterns evolve. Both studies focus on relabeling transducers, that is, transducers that map each individual input symbol to exactly one output symbol. This restriction simplifies the problem by ensuring a one-to-one alignment between the symbols in the input and output strings. The formal definitions of finite-state transducers studied in the thesis are recalled below as a basis for the problem formulation.

The set of natural numbers (excluding zero) is denoted by $\mathbb{N}$. For every $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, \ldots, n\} \subseteq \mathbb{N}$, and $[0, 1]$ denotes the unit interval

of real numbers. The number of elements of a finite set $X$ is written $|X|$. The powerset of $X$ is written $2^X$.

An *alphabet* $\Sigma$ is a finite nonempty set. $\Sigma^*$ denotes the set of all strings over $\Sigma$, including the empty string, denoted by $\varepsilon$. The length of a string $w \in \Sigma^*$ is written as $|w|$, and we denote by $w[i]$, where $i \in \mathbb{N}$ and $1 \leq i \leq |w|$, the $i$th symbol of $w$. Similarly, we write $w[i,j]$, $1 \leq i \leq j \leq |w|$, for the substring $w[i]w[i+1]\cdots w[j]$ of $w$.

**Definition 1** (Probabilistic Finite-State Transducer)**.** *A (relabeling probabilistic) finite-state string transducer (PFST) is a tuple* $M = (\mathcal{Q}, \Sigma, \Delta, \delta, q_0)$ *where* $\mathcal{Q}$ *is a finite set of* transducer states*;* $q_0 \in \mathcal{Q}$ *is the* initial transducer state*;* $\Sigma$ *and* $\Delta$ *are finite sets (called* alphabets*) of* input symbols *and* output symbols*, respectively. Finally,* $\delta\colon ((\Sigma \times \mathcal{Q}) \times (\Delta \times \mathcal{Q})) \to [0,1]$ *is a mapping such that for all* $\alpha \in \Sigma$ *and* $q \in \mathcal{Q}$*, it holds that*

$$\sum_{\beta \in \Delta, q' \in \mathcal{Q}} \delta((\alpha, q), (\beta, q')) = 1 \ .$$

*An element* $((\alpha, q), (\beta, q'))$ *of* $((\Sigma \times \mathcal{Q}) \times (\Delta \times \mathcal{Q}))$ *is called a* transition rule *and* $\delta((\alpha, q), (\beta, q'))$ *represents its probability. To denote that* $\delta((\alpha, q), (\beta, q')) = p$*, we write* $q \xrightarrow{\alpha/\beta,\, p} q'$*.*

A PFST is *(total) deterministic* if, for every $\alpha \in \Sigma$ and $q \in \mathcal{Q}$, there is exactly one $\beta \in \Delta$ and $q' \in \mathcal{Q}$ such that $\delta((\alpha, q), (\beta, q')) = 1$.

**Definition 2** (Finite-State Transducer)**.** *A (relabeling non-probabilistic) finite-state string transducer (FST) is a tuple* $M = (\mathcal{Q}, \Sigma, \Delta, \delta, q_0)$ *where* $(\mathcal{Q}, \Sigma, \Delta, q_0)$ *is a PFST (see Definition 1), and the output domain of* $\delta$ *is restricted to* $\{0, 1\}$*. When there is no risk of confusion, it is convenient to identify* $\delta$ *with the set of transition rules in* $((\Sigma \times \mathcal{Q}) \times (\Delta \times \mathcal{Q}))$ *that* $\delta$ *maps to the value* 1*. To further simplify notation, denote the fact that* $((\alpha, q), (\beta, q')) \in \delta$ *by writing* $q \xrightarrow{\alpha/\beta} q'$*.*

**Definition 3** (PFST semantics)**.** *Let* $M = (\mathcal{Q}, \Sigma, \Delta, \delta, q_0)$ *be a PFST. A* run *of* $M$ *on an* input string $w = \alpha_1 \cdots \alpha_n \in \Sigma^*$*, where* $\alpha_i \in \Sigma$ *for every* $i \in [n]$*, is a sequence*

$$\rho = q_0 \xrightarrow{\alpha_1/\beta_1,\, p_1} q_1 \xrightarrow{\alpha_2/\beta_2,\, p_2} q_2 \cdots \xrightarrow{\alpha_n/\beta_n,\, p_n} q_n$$

*of transition rules* $q_{i-1} \xrightarrow{\alpha_i/\beta_i,\, p_i} q_i$ *of* $M$ *(*$i \in [n]$*). The* output *of the run* $\rho$ *is* $outp(\rho) = \beta_1 \cdots \beta_n$ *and its* probability *is* $\delta(\rho) = \prod_{i=1}^{n} p_i$*. The set of all runs of* $M$ *on* $w$ *is denoted by* $runs_M(w)$*. Finally, the (probabilistic)* transduction computed by $M$ *is the mapping* $M\colon \Sigma^* \times \Delta^* \to [0,1]$ *such that, for all* $w \in \Sigma^*$ *and* $u \in \Delta^*$*,*

$$M(w, u) = \sum_{\substack{\rho \in runs_M(w) \\ outp(\rho) = u}} \delta(\rho) \ .$$

Definition 3 naturally specializes to the non-probabilistic case as defined in Definition 2. In such cases, each transition rule $q_{i-1} \xrightarrow{\alpha_i/\beta_i} q_i$ is assumed to have an implicit probability 1, so that for any run $\rho = q_0 \xrightarrow{\alpha_1/\beta_1} q_1 \cdots \xrightarrow{\alpha_n/\beta_n} q_n$ we have $\delta(\rho) = 1$. Consequently, for deterministic non-probabilistic FSTs, there exists exactly one output $u = M(w)$ for each input $w \in \Sigma^*$. Therefore, $M(w, u)$ acts as an indicator function taking value 1 if $u$ is the unique output produced by $M$ on $w$, and 0 otherwise. In this case, $M$ defines a total mapping $M : \Sigma^* \to \Delta^*$.

Figure 3.1 shows an example of a probabilistic finite-state transducer. This PFST operates on an input string in a symbol-by-symbol manner, producing exactly one output symbol for each input symbol. By reading the first input symbol $a$, it probabilistically selects one of two behaviors: with probability 0.5, it maps $a$ to $a$ and transitions to state $q_1$, and with probability 0.5, it maps $a$ to $b$ and transitions to state $q_2$. After this initial choice, the transducer follows a fixed behavior, consistently mapping all subsequent occurrences of $a$ to the same output symbol selected in the first step. In doing so, it defines a probabilistic mapping from input strings over $a$ to output strings over $a, b$, where each input string is mapped either to an identical string of $a$s or to a string of $b$s of the same length, each with probability 0.5.

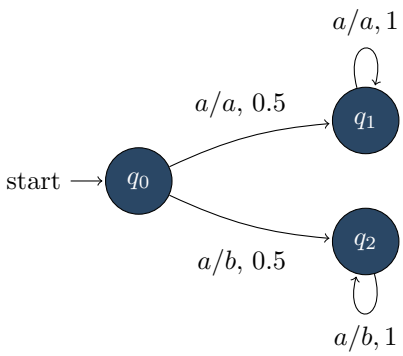



Figure 3.1: An example of probabilistic finite-state transducers.

These formal definitions provide the foundation for the grammatical inference methods presented in the following section, which treats the transducer state and transition rules of the transducers as learning objectives, inferred from observed input-output sequences.

## 3.2 Reinforcement Learning

Reinforcement learning is a major paradigm within machine learning and is widely applied to sequential decision-making problems. Unlike supervised or unsupervised learning methods, RL learns by continuously interacting with the problem environment and receiving a scalar reward signal that evaluates the quality of the actions taken (see Figure 3.2). One of the key properties of RL is its ability to handle delayed rewards, making it well-suited for problems where correct behavior emerges over a sequence of actions. This characteristic aligns naturally with grammatical inference of FSTs, where transduction behavior is realized through a sequence of symbol-level decisions and may only be evaluated after an entire output string has been produced.

Several considerations motivate the use of RL for learning finite-state transducers in this thesis. First, RL is inherently goal-oriented and adaptive. Its learning is always guided by an explicit objective encoded in the reward function rather than by fixed structural assumptions. This relaxes the limitation of traditional formal models and classical grammatical inference methods outlined in previous sections, which often rely on pre-defined structures or static rule sets. When interacting with an initially unknown transducer environment, an RL-based learner can gradually infer the latent mapping structure from observed input-output string pairs. Then, if the underlying transduction behavior changes over time, the learner can revise its inferred structure accordingly, allowing it to adapt to evolving contextual patterns rather than remaining bound to a fixed set of rules. In this sense, inferring a transducer structure and its transition rules provides a valid starting point, while the learning objective remains focused on achieving correct string mappings through continuous adaptation.

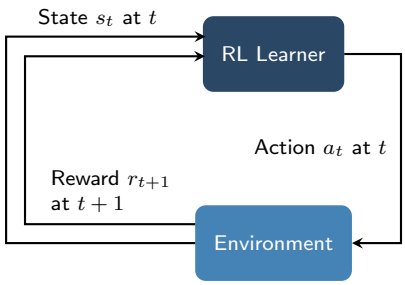



Figure 3.2: The reinforcement learning framework, which illustrates the interaction between the learner and the environment through states, actions, and rewards.

Second, an additional advantage of RL for grammatical inference lies in its explicit treatment of the exploration-exploitation trade-off. Learning transductions requires exploring alternative, potential mappings and transducer state transitions, while progressively exploiting those that have yielded good performance. This is particularly important in adaptive learning processing scenarios.

Third, there is a strong structural correspondence between FSTs and RL models. In an FST, the current transducer state encodes the relevant contextual information needed to determine how an input symbol is transformed. Similarly, in RL, the current state of the learner contains all relevant information about the environment required to select an action. Therefore, the combination of the current input symbol and the current transducer state can be directly formulated as the RL state, while producing an output symbol and transitioning to the next transducer state corresponds to taking an action in RL. This structural correspondence allows transduction learning to be naturally framed as a sequential decision-making problem.

Lastly, RL problems are commonly formalized as a *Markov Decision Process* (MDP), which assumes that future behavior depends only on the current state and the chosen action. This Markov assumption is consistent with the way transducers process input symbols, where the current state captures all necessary contextual information for subsequent processing. As a result, modeling grammatical inference within an MDP framework provides a structured way to connect formal transducer models with learning-based decision processes.

Within this framework, various RL methods can be employed depending on the properties of the target transducer. For deterministic, non-probabilistic transducers, tabular methods such as Q-learning [134] are suitable for their simplicity and interpretability. Q-learning estimates the value (i.e., the expected accumulated rewards) of state-action pairs and enables the learner to iteratively improve its transduction behavior based on observed rewards. Due to its explicit value representation, this approach preserves interpretability and aligns naturally with the discrete structure of FSTs. On the other hand, when learning probabilistic transducers, deep reinforcement learning methods extend the same decision-making principles with parameterization through neural networks. In this thesis, neural approximations are incorporated with Q-learning to estimate transition probabilities, while the underlying transducer structure and decision process remain explicitly defined. Therefore, the introduction of neural components supports probabilistic modeling without undermining the interpretability of the inferred translation. As the primary focus of this thesis lies in adopting RL to learn transducers rather than in algorithmic derivations, detailed mathematical definitions are provided in Papers I and II.

## 3.3 Keyword-Based Contextual Advertising

As the second major challenge explored in this thesis, Papers III and IV investigate contextual advertising with a particular focus on *keyword matching* as a concrete and practical instance of contextual language processing in real-world, user-centered applications. Within the broader landscape of contextual advertising, keyword matching represents a structured and widely deployed mechanism for associating textual content with advertising items. Although more complicated contextual matching approaches exist, keyword matching remains effective across various domains including online advertising and recommendation systems. Their continued relevance stems from their interpretability, domain-agnostic nature, and computational efficiency, which make them particularly suitable for privacy-sensitive and cold-start scenarios [31, 37].

Contextual advertising aims to place advertisements in situations (e.g., web pages) where they are most relevant to the surrounding content (for more details, see Section 2.2). Since both web pages and advertisements typically contain a lot of text, keywords and key phrases extracted from these texts provide a natural and compact representation of their semantic and contextual information [89]. By comparing the extracted keywords of a web page and an ad, contextual relevance can be estimated and used to guide ad placement decisions. These keyword representations enable efficient content matching, topical relevance inference, and audience targeting without relying on persistent user identifiers, thereby aligning with increasing privacy constraints.

Unlike traditional keyword targeting in sponsored search advertising, where advertisers explicitly bid on user-issued query terms [85], keyword-based contextual advertising operates in a broader and more automated setting. For

example, when placing an ad on a news article web page using keyword matching, relevance estimation relies on an automated pipeline in which *keyword extraction methods* play a crucial role. These methods aim to identify a small set of representative terms or phrases that summarize the main topics or salient concepts of a given document [38]. Existing approaches span a wide spectrum, ranging from simple statistical ones to more advanced neural models and large language models [52, 93, 122].

In this thesis, the general workflow of keyword-based contextual advertising in a news web page setting is conceptualized as a three-stage process. First, human experts assign keywords to ads, establishing a controlled representation of ad content. Second, keywords for news articles are automatically extracted using keyword extraction techniques. Third, ad-article relevance is measured by computing contextual similarity between the two keyword sets, typically based on lexical overlap or semantic similarity in a shared vector space [22]. This process results in a ranked list of candidate ad-article pairs. It defines a sequence of transformations from raw textual content to keyword representations, from keywords to similarity scores, and ultimately to targeting decisions.

This pipeline raises critical questions from both advertiser and user perspectives. For advertisers and advertising platforms, the effectiveness of keyword-based targeting directly influences campaign performance, relevance outcomes, and user engagement. While more advanced keyword extraction methods may produce more precise keywords, they also tend to require greater computational

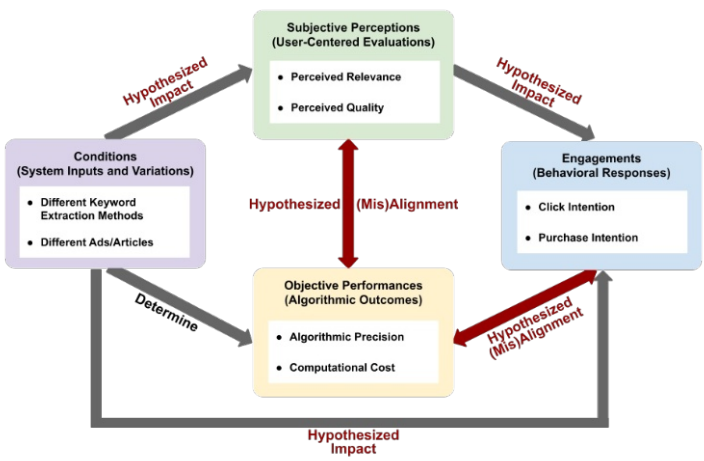



Figure 3.3: The conceptual model illustrates how system conditions, such as different keyword extraction methods, determine algorithmic performance outcomes, and how both algorithmic outputs and content characteristics relate to users' subjective perceptions and engagements.

resources and high maintenance costs. However, whether these additional costs translate into meaningful gains in user engagement and perceived relevance remains unclear. From the user perspective, keyword-based targeting shapes how advertisements are experienced during everyday content consumption, which affects whether ads are perceived as relevant, informative or intrusive. Despite keyword extraction being a well-studied NLP task with numerous established methods, it remains unclear what constitutes good keywords from the perspective of real users encountering news articles in everyday contexts.

Accordingly, the core problem addressed in Papers III and IV concerns the potential divergence between algorithmic relevance, optimized by keyword extraction and similarity computation, and human-perceived relevance, as well as the downstream consequences of this divergence for user engagement behaviors. To structure this problem formally, Figure 3.3 presents a unified conceptual model that integrates keyword extraction methods, computational resource considerations, and user response based on the theoretical foundations discussed in Section 2.2. This model provides the analytical foundation for the empirical investigations in Papers III and IV, conducted using the methodologies introduced in the following sections. Paper V further extends these discussions by conceptualizing these trade-offs at a higher level in a broader perspective, focusing on how technical, institutional, and perceptual factors jointly influence online advertising strategies.

## 3.4 User Experiments

Evaluating algorithmic performance in contextual language processing, particularly in NLP tasks, has largely relied on benchmarking against *gold-standard* datasets. These gold standards typically consist of manually annotated corpora that represent high-quality target outputs for specific domains or tasks, and they are widely used for both training and evaluation purposes [136]. In keyword extraction, gold standards are intended to provide reference annotations for assessing the accuracy of automatic extraction methods [56], often measured by metrics including F1-score and cosine similarity. Well-known benchmark datasets include SemEval-2010, KP20k and Inspec [13, 62, 96].

While this evaluation paradigm remains fundamental for measuring algorithmic accuracy, it faces increasingly practical and conceptual limitations in applied, user-centered domains. Suitable and openly available gold-standard datasets are often scarce, and many existing benchmarks may become outdated or lack transparency regarding their annotation origins and quality [84, 90, 119]. More importantly, precision-oriented benchmarking focuses primarily on surface-level algorithmic outputs, offering limited insight into how these outputs are perceived or experienced by end users. In domains such as contextual advertising, where its effectiveness ultimately depends on user interpretation and response, the alignment between computational performance and human understanding becomes a critical concern.

To address these limitations, Papers III and IV combine traditional algorithmic evaluation with user-centered assessment through online survey experiments. This combined approach is designed to bridge algorithmic and perceptual perspectives, to enable a more comprehensive evaluation of contextual language processing methods in practice. In addition, online surveys are a popular methodology in advertising and marketing research, as they allow access to a wide range of users and capture their first-hand subjective feedback [26, 64]. Online surveys offer flexibility in their design, enabling precise control over stimuli presentation, question formats and response collection with respect to the specific experimental goals.

To conduct such survey experiments, both studies adopt a unified framework to empirically examine how different keyword extraction methods influence user perceptions in a keyword-based contextual advertising scenario. This framework is structured into four main stages: conceptualization, preprocessing, implementation, and data analysis. The conceptualization stage involves theoretical preparation guided by the conceptual model presented in Figure 3.3, including the formulation of key research questions, the design of experimental manipulations and factorial structures, the development of survey instruments and sample power analysis. In contrast, the pre-processing stage focuses on technical preparation derived from conceptual decisions, such as generating experimental stimuli and preparing data materials. For instance, in Paper IV, three algorithmic keyword extraction methods are selected as part of the experimental manipulation and their corresponding keyword outputs are generated during this stage. The implementation stage covers the deployment and execution of the online experiments. Finally, the data analysis stage cleans and statistically analyzes (see Section 3.5) the collected user responses to support interpretation and discussion.

During the implementation stage, annotator participants are recruited through the crowdsourcing platform *Prolific*[1]. Prolific allows researchers to specify detailed screening criteria such as location, age and gender, to ensure that the sample aligns with the experiment objectives. It also provides a set of tools that help streamline participant management and facilitate integration with external survey or annotation platforms. For survey design and data collection, the web-based annotation tool *Potato*[2] [108] is used. It offers customizable templates for building online survey interfaces, including Likert-scale questions and other interactive components. Its automated pipelines support real-time monitoring and structured outputs, thereby increasing the reliability and efficiency of both the survey experiment process and subsequent data analysis. To further support reproducibility and encourage the adoption of user-centered evaluation practices, the full procedure for integrating *Potato* with *Prolific* has been documented and released in a public GitHub repository[3], hoping to help others who may otherwise face similar challenges in setting up such experiments.

[1]See `https://www.prolific.com`

[2]See `https://potato-annotation.readthedocs.io/en/latest/`

[3]See `https://github.com/JingWen17/prolific-potato-pipeline.git`

## 3.5 Statistical Analysis Tools

Statistical analysis provides a systematic means to quantify perception-based user ratings. It enables testing of validity, consistency and statistical significance of the collected user responses. In addition, it helps identify potential causal relationships among the measured factors, thereby clarifying causal effects and providing a solid basis for drawing conclusions and implications.

A wide range of well-established statistical methods are available for this purpose, including the t-test, p-test, Chi-square test, and Kendall's W [40, 71, 74, 81]. Kendall's W is a widely used nonparametric statistic for assessing inter-rater reliability and consistency, particularly suitable for ordinal ratings. However, its general use requires that each experimental task be evaluated the same number of times, which is not the case in the experimental design of Paper III. Therefore, Paper III develops two statistical methods to evaluate the reliability and consistency of user ratings across different survey question phrasings, from both the experimental task and participant perspectives, under the assumption that the most effective survey question phrasing should yield consistent and less variable user ratings. To further quantify the influence of experimental manipulation factors on user perceptions, the correlation-based analyses, such as the Chi-square test and the t-test, are used. These tests examine the strength and direction of relationships between variables and assess whether observed differences in user ratings are statistically significant.

Paper IV adopts a more comprehensive and systematic statistical framework to evaluate the structure of the experimental design and to uncover the underlying causal relationships between experimental manipulations and measured factors. A *Confirmatory Factor Analysis* (CFA) [15] is first conducted to evaluate whether the survey items effectively measured their intended latent constructs. CFA also inspects inter-factor correlations, assesses discriminant validity, and refines the theoretical structure for subsequent modeling. Figure 3.4 shows an example of a two-factor, six-indicator CFA model. Circles denote latent variables, whereas squares represent manifest or observed variables. Single-headed arrows indicate directional relationships such as factor loadings from latent factors to observed indicators, denoted by $\lambda$ in the figure. Double-headed arrows represent covariances or correlations between latent variables, denoted by $\Psi$. The $\varepsilon$ in the figure corresponds to the residual or error variances associated with each observed variable. In this model, six observed variables are hypothesized to load onto two latent factors with vary-

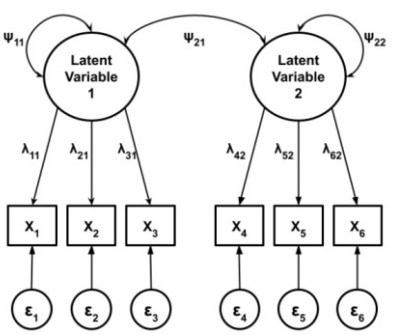



Figure 3.4: An example of a two-factor CFA measurement model with six manifest indicators.

ing strengths, as reflected by their respective factor loading. It statistically evaluates whether the observed data are consistent with this predefined factor structure by estimating the loadings, residual variances, and latent factor correlations. The covariance between the two latent factors captures the extent to which they are statistically related.

Building on these results, Paper IV then applies a *Structural Equation Modeling* (SEM) [130] approach to construct and test the hypothesized causal pathways between the manipulations and measured factors. A detailed illustration of an SEM model is provided in Section 4.4. SEM not only quantifies the strength of individual causal effects but also is capable of capturing complex, indirect, and cross-factor relationships. Together, CFA and SEM provide a robust analytical framework for testing theoretical assumptions, validating measurement models, and uncovering deeper insights into the mechanisms underlying user perceptions and behaviors.

# Chapter 4

# Summary of Contributions

This thesis investigates contextual language processing from both methodological and applied perspectives. It comprises five papers that span two major research directions: (i) the design of adaptive learning methods for grammatical inference of finite-state transducers, and (ii) the exploration of key factors influencing the effectiveness of contextual language processing in online contextual advertising. Together, the thesis examines how contextual information can be learned and evaluated in both formal language models and real-world, user-facing applications. Specifically, it addresses the following research aims:

> **AIM 1.** Advance contextual language processing through interpretable, learning-based models and user-centered evaluation in AI-driven applications such as contextual advertising.
>
> **AIM 2.** Strengthen the theoretical and methodological foundations of contextual language processing by designing adaptive learning algorithms.
>
> **AIM 3.** Bridge algorithmic and perceptual perspectives in contextual language processing to better align computational performance with human understanding, considering constraints and implications for AI-driven advertising practices.

Papers I and II address AIM 1 and AIM 2, in which we incorporate reinforcement learning-based frameworks for the grammatical inference of finite-state transducers, a foundational formal model for contextual language processing. By formalizing transducer learning as an adaptive sequential decision-making problem, these studies demonstrate how structured language models can be learned from data while preserving interpretability and enabling adaptation to evolving contextual patterns. In doing so, they advance both the theoretical and methodological foundations of adaptive contextual language processing.

Papers III and IV mainly contribute to addressing AIM 1 and AIM 3 by examining contextual language processing in the applied setting of keyword-

based contextual advertising. These papers emphasize the importance of user-centered evaluation and propose a combined assessment framework that integrates traditional algorithmic benchmarking with controlled user survey experiments. Through statistical analysis, the studies investigate how different keyword extraction methods influence user perceptions of relevance, trust, and effectiveness, highlighting an under-explored gap between algorithmic performance and user experience. The findings provide practical insights into how contextual language processing methods should be selected and evaluated in real-world advertising environments.

Paper V further contributes to AIM 3 by extending the discussion beyond keyword-based contextual advertising to the broader domain of programmatic advertising. The paper builds a conceptual model to characterize AI-driven advertising practices in terms of resources, rectitude, and results. It highlights the interplay between algorithmic capabilities, institutional constraints, and user perceptions, and offers explicit recommendations for effective and responsible advertising strategies that balance multiple practical and ethical considerations.

**Author contributions.** The author of this thesis is the main contributor to Papers I-IV. Specifically, in Papers I and II, she designed and conducted all experiments, prepared the visual presentation of the findings, and contributed to idea conception and manuscript writing. In Papers III and IV, she was the lead contributor, responsible for idea conception, experimental design, implementation, data collection and analysis, visualization, and manuscript writing. Paper V is a collaborative work involving seven authors, where she primarily contributed to the idea conception and manuscript writing as a co-author.

## 4.1 Paper I

**Reinforcement Learning of Finite-State String Transductions** (Johanna Björklund, Jingwen Cai and Anna Jonsson, 2024)

*Paper I uses a reinforcement learning-based algorithm and proposes a state-splitting method for the grammatical inference of deterministic, non-probabilistic string-to-string finite-state transducers.*

In this paper, the grammatical inference problem for finite-state string transductions is formulated as a sequential decision-making task. Rather than treating transducer inference as a purely symbolic problem, the paper models the transduction process as a finite Markov decision process, in which decisions are made incrementally at the symbol level. At each timestep, the RL learner observes the current environment state, which includes the current transducer state and the current input symbol, and selects an action that jointly determines the output symbol and the next transducer state. Reward-based feedback, provided either instantly or in a delayed manner, reflects the quality of the generated output string relative to the target transduction. This formulation enables the application of tabular Q-learning to learn non-probabilistic transduction behavior directly from input-output string pairs. Through re-

peated interaction with the environment, the learner incrementally improves its strategy, effectively learning both the output mapping and the underlying state-transition structure of the transducer.

A key methodological contribution of the paper is the proposed state-splitting RL algorithm, which eases the reliance on a fixed, predefined number of transducer states before learning. Instead of assuming the target transducer structure in advance, the learner begins with a single transducer state and dynamically expands its state and action space during learning. When a state consistently yields low rewards, this may indicate that it is confused by multiple incompatible contextual situations. Therefore, the algorithm splits this state into two, allowing the learner to better separate distinct syntactic or contextual classes. The underlying idea is that if there is no good choice of actions, then it is likely because the state is reached by strings that represent different syntactic classes with respect to the target transduction, so there is not one appropriate way forward. This mechanism enables the transducer structure to emerge adaptively from interaction with data, guided by the learning objective rather than by prior structural assumptions.

The effectiveness of the proposed framework is evaluated through a series of experiments. In each experiment, the RL learner is set to learn a transducer such as *IDENTITY* (which computes an output string equal to the input string), *B-TO-A* (every $b$ in the input string is replaced by an $a$ in the output string), and *n-COUNT* (every $n$th input symbol is replaced by a $c$ in the output string), through input-output strings sampled by *pyfoma* [61]. Although the target transductions used in these experiments are conceptually simple, learning them from the defined reward signal alone is not trivial. Specifically, under the delayed reward setting, which better reflects realistic interaction scenarios, feedback is sparse and only provided after processing the entire string, resulting in a weak reward signal and a challenging learning problem. For example, when learning the *B-TO-A* transduction with the sample input string $ab$ at the beginning, the learner receives no feedback before processing the entire string, and all actions initially appear equally valuable. At the end, the learner may receive a negative reward indicating a single mismatched symbol. However, this feedback alone does not reveal which decision led to the error, since generating $ab$ and $ba$ ($aa$ is the target output string in this case) yields the same reward but corresponds to fundamentally different transduction behaviors. This challenge becomes more pronounced as the length of the input strings increases, as the number of transducer states grows, and as the action space expands with additional symbols. In such cases, an incorrect output symbol may result in only a small penalty relative to the total string length, making it difficult for the learner to immediately identify the specific decision responsible for the error. Despite these challenges, continued interaction and exploration help the learner gradually resolve these ambiguities and converge to the correct transduction. Therefore, the experimental results show that even with weak feedback, the proposed framework can efficiently infer correct transduction behavior and represent an interpretable transducer structure.

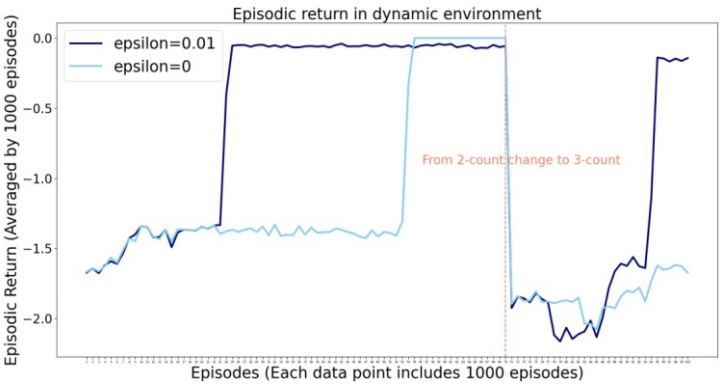



Figure 4.1: The episodic returns obtained by the learner in the dynamic environment. The higher the better output strings generated by the learner. The target transduction is initially *2-COUNT*, but at episode 70 000, it is replaced by the transduction *3-COUNT*.

Beyond static learning scenarios, the paper further investigates learning in a dynamic environment, where the target transduction changes during training. Figure 4.1 shows the learning results in such a dynamic environment. In this learning case, the initial target transduction is replacing every second letter in the input string by a *c*, but during the learning, the target transduction changes to replace every third letter by a *c*. Results show that, when combined with the epsilon-greedy exploration method, the RL learner can adapt to changes by revising its inferred structure and decision-making policy, rather than remaining bound to previously learned rules. By demonstrating such revising behaviors of the learner, this paper helps to picture a way of how adaptive learning provides a mechanism for modeling contextual variation. Rather than relearning from scratch or requiring manual re-specification, the learner incrementally updates its strategy and structure in response to new evidence.

To the best of our knowledge and at the time of publication, this paper represents the first attempt to apply reinforcement learning to the inference of FSTs. Although the study is restricted to deterministic, non-probabilistic transducers and equal-length input-output strings, it establishes a foundational RL-based formulation of FST inference and provides an experimental basis for evaluating its effectiveness. These contributions address AIM 1 and AIM 2 by demonstrating how structured language transformations can be learned adaptively from data while preserving interpretability. The proposed learning formulation and the state-splitting method strengthen the methodological foundations of contextual language processing and illustrate how classical formal models can be extended to adaptive, learning-based approaches.

## 4.2 Paper II

**Reinforcement Learning of Probabilistic Finite-State Transducers** (Jingwen Cai, Johanna Björklund and Frank Drewes, 2025)

*Paper II introduces a two-phase iterative framework that builds upon the table-based reinforcement learning, deep reinforcement learning, and the state-splitting methods for learning probabilistic finite-state string transducers.*

Building on the deterministic formulation introduced in Paper I, this paper addresses the grammatical inference of probabilistic finite-state transducers, where each transducer state and input symbol may have multiple valid output transitions with associated probabilities. While the learning problem is again formulated as a Markov decision process, the probabilistic setting introduces additional complexity in three aspects: first, the learning goal is not only to infer an appropriate transducer structure (i.e., identify the transducer states and the corresponding transductions), but also to approximate the probability distribution over alternative transitions; second, previous reward formulation based on output mismatches is insufficient for this purpose, because RL tends to converge to a single optimal action even when multiple actions are equally optimal under the target distribution. In addition, this mismatch reward is not able to inform the learner how accurate the probability approximation is; lastly, table-based Q-learning is not sufficient in learning PFSTs, so more powerful tools are needed for inferring the probabilities.

To address these challenges, the paper proposes a two-phased iterative framework (see Figure 4.2) that separates structure inference from probability approximation. In Phase I, the goal is to infer a reference transducer structure. This phase adopts tabular Q-learning and the state-splitting mechanism introduced in Paper I, starting from a single transducer state and incrementally expanding the structure through interaction with sampled input-output string pairs. To handle probabilistic behavior, the PFST is abstracted as a set of deterministic transducers, each corresponding to a possible realization of the probabilistic process. The learner then interacts with these realizations independently, allowing it to aggregate the resulting experiences into a set of optimal actions associated with each transducer state across realizations, without collapsing prematurely to a single behavior. This decompose-compose strat-

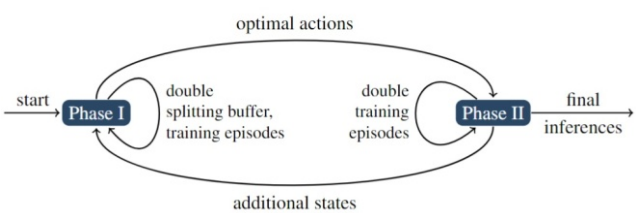



Figure 4.2: The two-phased iterative framework

egy helps preserve alternative valid transitions during interpretable structure learning and provides a solid foundation for subsequent probability estimation.

Once a sufficiently expressive structure and action set are identified, Phase II focuses on learning the transition probabilities. This phase employs a deep reinforcement learning method based on Soft Actor-Critic (SAC) [54], which incorporates entropy regularization to encourage stochastic policies. This property is essential for approximating probability distributions rather than deterministic mappings. The learning objective is defined in terms of minimizing the Kullback-Leibler (KL) [79] divergence between the output distributions induced by the learned transductions and that of the target PFST. Before transitioning between phases, each phase may iterate internally with increasing the splitting buffer, which serves as a cool-down period that allows sufficient exploration before deciding to split states, and extending the number of training episodes if the phase-level performance fails to meet a performance threshold. By iteratively alternating between structure refinement and probability estimation, the framework ensures that probability learning is grounded in an adequate structural representation.

The experimental results further validate the effectiveness of the proposed two-phase learning framework. They demonstrate that the learner is able to correctly infer both the structure and the associated transition probabilities of the target PFST, starting from minimal prior assumptions about the transducer architecture. In particular, the results show that separating structure inference from probability estimation leads to more stable convergence than attempting to learn both aspects jointly with either of the phases alone.

By extending RL-based grammatical inference to probabilistic finite-state transducers, this paper demonstrates how transduction uncertainty and alternative transitions can be learned. The proposed framework incorporates deep RL to capture probabilities where simpler methods struggle with, while retaining an interpretable, state-based transducer structure learning process. It emphasizes how such learning can enrich classical formal models to better reflect the inherent uncertainty and multiplicity of contextual language phenomena, thereby contributing to addressing AIM 1 and AIM 2.

## 4.3 Paper III

**From Precision to Perception: Human-In-The-Loop Evaluation of Keyword Extraction for Internet-Scale Contextual Advertising** (Jingwen Cai, Sara Leckner and Johanna Björklund, 2026)

*Paper III highlights the long-overlooked gap between traditional precision-based algorithmic evaluation and real user perceptions in real-world applications through four online user experiments across six keyword extraction methods, including large language models.*

In Paper III, four online, survey-based experiments are designed and conducted to investigate user perceptions of keyword extraction methods applied

to five news articles and their corresponding downstream tasks, following the experimental procedure outlined in Section 3.4. Prior to these experiments, ten keywords are extracted for each article using TF-IDF, KeyBERT, and Llama 2 [52, 122, 126], representing increasing levels of model complexity. The experiments examine four key properties that characterize high-quality keywords: comprehensiveness, representativeness, distinctiveness, and reasonableness. In total, ratings from 855 participants are analyzed in the paper. Participants are recruited through Prolific and complete the online surveys designed and implemented using Potato.

The studies consist of four stages. A preliminary experiment is first conducted to validate the online survey procedure and identify the most consistent phrasings of survey questions for each property. To this end, the paper introduces two complementary analytical approaches for assessing the consistency of these phrasings on participant ratings: the *horizontal* and *vertical analysis*, which evaluate across survey instances and participants, respectively. Gold-standard keyword sets are also derived in this experiment based on participant-selected keywords, serving as the benchmark for evaluating algorithmically generated keywords in subsequent experiments. The main experiment collects participant ratings for each article-method combination, including the gold-standard keywords, using the finalized survey questions verified from the preliminary experiment. A supplementary experiment provides an additional perspective by asking participants to rank all keyword sets generated for the same article by all four methods (i.e., TF-IDF, KeyBERT, Llama 2, and Gold Standard) according to their perceived qualities. Finally, a downstream experiment simulates a real-world article recommendation scenario. For each base article, five most related news articles are retrieved based on keyword similarity using an extended set of keyword extraction methods, including three additional advanced LLMs: Llama 3, Mistral, and DeepSeek [28, 47, 69]. Participants then evaluate the perceived relevance of the resulting recommendations.

Analysis of participant ratings across all experiments provides empirical insights into how users perceive the quality of keywords generated by different methods. The results demonstrate the importance of incorporating user perceptions in algorithmic evaluations. Specifically, participant ratings for keywords extracted by TF-IDF, KeyBERT, and Llama 2 do not differ significantly, but all are rated with notably lower scores compared to the participant-selected gold-standard keywords. This highlights the necessity of including human judgment in keyword-based applications. Results from the downstream experiment (see Figure 4.3) further indicate that Mistral significantly outperforms both simpler models and other LLMs, even the much larger-scaled DeepSeek, in article recommendation tasks in terms of user perceptions. Interestingly, although DeepSeek achieves the highest precision-based scores in traditional benchmarking against gold standards, it does not perform significantly better than KeyBERT in user evaluations, as shown in Figure 4.3. These findings highlight the efficiency of lightweight models in user-oriented tasks and underscore a critical gap between precision-oriented algorithmic assessment and user-centered eval-

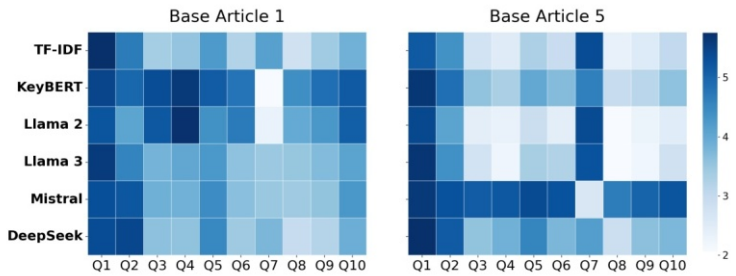



Figure 4.3: Averaged evaluation scores for Base Articles 1 and 5 in the downstream experiment: In each heat map, rows represent task questions, and columns represent keyword extraction methods. Cell colors indicate the average ratings for each question-method pair, with darker shades reflecting higher scores (reversed for Q7).

uation, and suggest that models optimized for technical metrics may not align with real user preferences.

In addition, the paper examines the practical implications of deploying different models by comparing the computational costs of using TF-IDF, KeyBERT, and Llama 2 for identical extraction tasks on a CPU. The results show that Llama 2 requires substantially more computation time than the simpler methods. Although GPUs and cloud-based services can reduce processing time for LLMs, similar improvements also apply to lightweight methods, leaving relative efficiency differences largely unchanged. Moreover, the expertise and financial resources required to deploy and maintain such facilities must be considered when evaluating their efficiency and real-world applicability.

Paper III contributes to AIM 1 by advancing user-centered evaluation practices for contextual language processing and to AIM 3 by empirically demonstrating the misalignment between algorithmic optimization and human perception. While the primary contribution of this paper lies in its user-centered evaluation of keyword extraction methods, the included models serve mainly as representative examples to illustrate a broader argument: beyond pursuing algorithmic precision and scaling model complexity, it is essential to consider what real users truly value in practice. Given the rapid evolution of LLMs and their increasing integration into everyday life, user-centered evaluation is critical for aligning algorithmic development with human perception, especially in real-world applications whose ultimate goal is to meet user needs, such as contextual advertising and recommendation systems.

## 4.4 Paper IV

**Beyond Precision: Understanding the Impact of Algorithmic Accuracy and Transparency on User Perceptions in Keyword-Driven Contextual Advertising** (Jingwen Cai, Bart Piet Knijnenburg, Johanna Björklund and Sara Leckner, 2026)

*Paper IV investigates user perceptions in keyword-based contextual advertising recommendations through an online user experiment comparing three keyword extraction methods, including an LLM, analyzed by systematic factor analysis methods.*

Building on the perceptual insights from Paper III, Paper IV presents a systematic analysis of user perceptions and behaviors through a more practical keyword-based contextual advertising experiment. The study focuses on keyword-based ad placement and examines not only user-perceived relevance, but also their behavioral intentions.

The experimental setup simulates a realistic advertising scenario. Five fictitious advertisements are created, and three domain experts manually select representative keywords that best describe each ad. News articles are drawn from the open-source *KPTimes* dataset, and article keywords are automatically extracted using three representative methods with increasing levels of model complexity: TF-IDF, KeyBERT, and DeepSeek. For each ad, the five most relevant articles are selected based on keyword similarity. Two baseline conditions are included for comparison with the algorithmic methods: a Gold Standard condition using the pre-annotated keywords provided by the dataset, and a Random condition in which news articles are assigned to each ad arbitrarily. A total of 498 participants recruited from Prolific evaluate the resulting ad-article pairs.

Compared to Paper III, Paper IV expands the scope of evaluation by incorporating behavioral intentions alongside perceptual judgments. Participants rate perceived relevance and congruence between ads and articles, as well as their intention to click on and purchase from the advertised products. These behavioral measures enable a deeper and more practical understanding of how different keyword extraction methods influence both user perceptions and behavioral intentions. In addition, the paper examines the role of ad transparency by manipulating the availability of explanations on the ad placement. Specifically, half of the participants are shown the keywords used to generate recommendations, while the other half receive no information about how the ad-article pairs are matched.

To model the relationships among these constructs, the paper employs Confirmatory Factor Analysis followed by Structural Equation Modeling. The CFA results reveal strong correlations between perceived relevance and congruence, as well as between click and purchase intention, leading to their consolidation into two higher-level constructs: perceived relevance and behavioral intention. The resulting SEM model further captures both direct and indirect effects among keyword extraction methods, explanations, perceived relevance, behav-

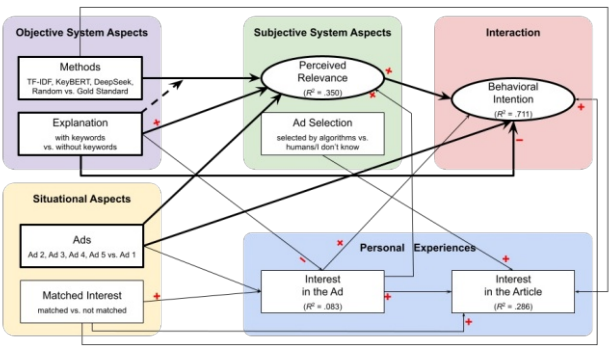
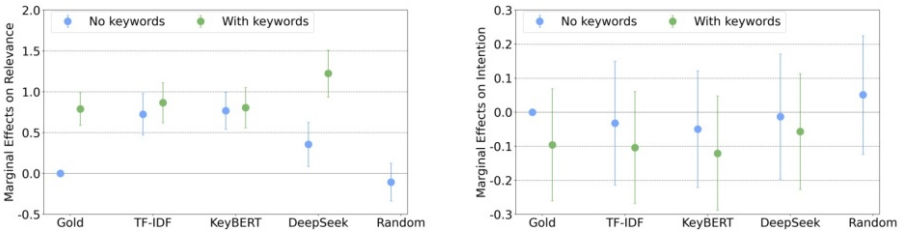



Figure 4.4: The structural equation model for the experimental data. Arrows represent causal effects, with "+" indicating positive effects and "-" the negative effects compared to the baselines. $R^2$ is the proportion of variance explained by the model.

ioral intention, and other influential factors. The key components of the SEM model are highlighted in Figure 4.4.

The SEM analysis reveals significant causal relationships among the measured factors. Results indicate that different keyword extraction methods influence participants' perceived relevance to varying degrees, with simpler methods such as TF-IDF and KeyBERT outperforming DeepSeek when explanations are absent (see Figure 4.5). When keywords are shown, however, DeepSeek achieves the highest perceived relevance, reflecting the higher semantic and reasonable quality of its extracted keywords. In addition, perceived relevance positively impacts user behavioral intention, confirming its central role in contextual advertising effectiveness. Crucially, providing explanations produces a paradoxical effect. While explanations increase perceived relevance, they simultaneously reduce behavioral intention. This negative impact is mediated primarily through diminished ad interest, which emerges as the strongest predictor of click and purchase intention in the model. These direct and indirect effects are sufficiently strong that the mediating role of perceived relevance in positively linking explanations to intentions becomes secondary.

These findings can be interpreted through established psychological frameworks. From the perspective of the Persuasion Knowledge Model, explicit explanations may activate users' awareness of persuasive intent, leading to resistance and reduced engagement. Similarly, the *Dual Process Theory* [41] suggests that while explanations support more deliberate, analytic processing of relevance, they may disrupt affective or heuristic pathways that drive spontaneous interest and reaction. Thus, greater transparency does not uniformly improve advertising efficiency and may backfire depending on how such contextual, explanatory information is conveyed.




(a) (b)

Figure 4.5: Marginal effects of keyword extraction methods and explanations on (a) perceived relevance and (b) behavioral intention. The effect of Gold Standard without showing keywords as explanations on relevance is set to zero, and the error bars indicate 95% confidence interval.

Beyond its empirical contributions, Paper IV offers important practical implications for users, researchers, and advertisers. It argues that increasing algorithmic sophistication or transparency does not automatically translate into improved user perception and engagement. Lightweight methods can achieve competitive performance when evaluated from a user-centered perspective, especially under realistic resource and deployment constraints. Moreover, it points to the need for future interdisciplinary research into different explanation designs and their downstream effects on user engagement, as poorly designed transparency can undermine user trust and effectiveness. Overall, Paper IV contributes to AIM 1 and AIM 3 by empirically and conceptually demonstrating how algorithmic relevance, explanation mechanisms, and human perception and engagement interact. By bridging computational modeling with psychological theory and empirical evidence, the paper highlights the necessity of aligning algorithmic design choices with human perceptions in AI-driven advertising applications.

## 4.5 Paper V

**Programmatic Advertising in the Age of AI: A Conceptual Overview and Strategic Recommendations** (Adam Åbonde Garke, Richard Wahlund, Sara Leckner, Emil Häglund, Jingwen Cai, Igor Ryazanov, and Johanna Björklund, 2025)

*Paper V proposes a conceptual framework for AI-driven online advertising, identifies key factors influencing advertising strategies, and offers correspondingly practical guidance for practitioners.*

Paper V examines how recent developments in artificial intelligence and contextual language processing are transforming the landscape of online programmatic advertising. The analysis identified two major shifts in the field: the evolution from personalized to contextual targeting, and the increasing integra-

tion of AI-driven automation. Data protection regulations such as the GDPR have compelled marketers to adopt greater transparency in their practices. In response to these stricter regulations, contextual advertising has emerged as an attractive alternative, as it relies primarily on the content of web pages and advertisements rather than on user personal data. At the same time, the rapid advancement of AI technologies continues to reshape advertising operations. These technologies not only enhance efficiency but also expand the scope and precision of digital marketing activities.

Building on these two trends, the paper proposes a framework encompassing six advertising strategies, which span the spectrum from contextual to personalized targeting and from AI-based to rule-based automation. Each strategy represents a unique combination of these two dimensions. For example, contextual targeting using AI-based automation involves placing ads through advanced contextual analysis and dynamic optimization powered by AI. In contrast, personalized targeting using rule-based automation refers to marketers deciding to whom the ads should be shown.

Paper V further identifies three critical factors that influence the selection and performance of these advertising strategies: results, resources, and rectitude. Results are the outcomes of a campaign in achieving its marketing objectives. For example, while contextual targeting supported by AI can help reach new audiences efficiently, it can be challenging for it to re-engage inactive customers. Resources are concerned with what capabilities and assets are available to marketers. Advanced AI-driven strategies typically demand greater computational power, technical proficiency, and financial investment compared to simpler rule-based approaches. Rectitude involves the ethical, legal, and regulatory dimensions of advertising practices. While potentially more precise, personalized targeting strategies often raise significant concerns about user privacy, regulatory compliance, and the reliance on third-party data.

Synthesizing these strategies and factors, the paper argues that while AI significantly enhances the scalability and efficiency of modern online advertising, it also introduces critical trade-offs. Increased AI-driven automation can reduce human oversight, creativity, and contextual sensitivity, and may heighten the risk of algorithmic bias or unintended outcomes. In addition, similar trade-offs exist between contextual and personalized targeting. Although personalized targeting can deliver highly tailored ads, it may also evoke feelings of exposure or invasion of privacy among users.

Paper V contributes to AIM 1 and AIM 3 by extending contextual language processing from algorithmic development to a strategic perspective on AI-driven advertising. By considering these trade-offs and the interplay of results, resources, and rectitude, marketers can make more informed and context-sensitive decisions about which advertising strategies to employ. Understanding these dynamics allows advertisers to balance efficiency, ethics, and effectiveness in the evolving ecosystem of AI-driven online advertising.

# Chapter 5

# Dicussion

This chapter synthesizes the key findings of the papers included in this thesis along two complementary dimensions. First, it examines the results from the perspective of learning mechanisms in fundamental models for contextual language processing. Second, it considers contextual language processing from a practical standpoint, focusing on its implications for real-world applications, with contextual advertising serving as a representative use case. The chapter then concludes with a discussion of directions for future research motivated by the insights and limitations identified in this work.

## 5.1 Interpretable and Adaptive Contextual Decision-Making

Paper I and Paper II demonstrate the feasibility of applying reinforcement learning to the grammatical inference of string finite-state transducers from sample string pairs. Rather than framing grammatical inference solely as a structural reconstruction problem, the papers emphasize a more interaction-oriented view of contextual language processing, in which learning is formulated as a sequential decision-making process. Under this perspective, contextual information is implicitly encoded in transducer states and transitions, while the learning objective is realized through explicit, step-by-step mapping decisions, guided by the reward signals alongside the interaction.

Within this formulation, RL provides an effective mechanism for discovering implicit structure and contextual dependencies through interaction and feedback. Specifically, transduction decisions are formalized in terms of state-action-reward paths, where the reward construction guides learning outcomes and reflects task-specific objectives. This framework establishes a correspondence between FSTs and RL algorithms by collapsing their central components, states, transition actions, and reward-guided objectives, into a unified theory-decision representation. As a result, the underlying language transformation

model and its contextual dependencies can be learned adaptively, rather than being fixed and pre-defined.

This adaptivity is particularly crucial for real-world applications, where contextual information and language usage may evolve over time. RL enables continuous adjustment of transduction behavior in response to changing environments, which is beyond what a fixed or pre-defined transducer model can express. In this sense, the learning process can be interpreted as training a "student" transducer to approximate and refine the behavior of an "expert" transducer, while remaining responsive to newly observed contextual language samples. Consequently, contextual properties and their inference emerge as intrinsic and dynamic aspects of the interactions and the learned model, rather than as static, manually specified rule sets.

In addition, the proposed state-splitting approach used in both papers further supports adaptive learning. Instead of pre-specifying the exact number of transducer states, the learning process begins from a minimal configuration and incrementally increases model capacity through state splitting. This approach allows the transducer structure to be constructed recursively from observed data, providing greater flexibility than methods that rely on a fixed or arbitrarily chosen number of states. Moreover, starting from a single state constrains the initial state and action spaces, thereby improving efficiency and reducing the risk of misleading or unstable exploration in the early stages.

Across both Q-learning and Soft Actor-Critic approaches explored in Papers I and II, the resulting models retain a relatively high degree of interpretability. Learned state values and strategies provide explicit representations of contextual dependencies and transition preferences. In particular, the tabular Q-learning method offers a straightforward, simple look-up explanation of the learned transduction outcomes. Compared with other neural network-based approaches, such representations substantially enhance the interpretability and traceability of decision processes.

In practical, user-facing scenarios, this interpretability supports explainability and transparency in contextual decision-making. Furthermore, the reward-driven nature of RL enables flexible, task-oriented optimization, allowing contextual relevance to be encoded implicitly through learning objectives rather than hard-coded rules. Taken together, the combination of finite-state transducers and reinforcement learning exemplifies an interpretable and adaptive approach to contextual decision-making, occupying a middle ground between rigid rule-based and opaque large-scale neural models.

## 5.2 Aligning Computational and User-Centered Contextual Relevance

Papers III, IV, and V examine contextual language processing from an applied perspective, with a particular focus on contextual advertising. Collectively, the findings demonstrate that contextual communication in real-world systems is

not determined solely by algorithmic design, but also relies on the interaction between users, computational mechanisms, and institutional constraints. Contextual relevance, therefore, should be understood as a socio-technical construct shaped by technical capabilities, user perception, and regulatory or ethical considerations, rather than as a fixed or universal solution.

One of the central findings from both Paper III and Paper IV is that system-level relevance does not necessarily coincide with user-perceived relevance, and that user behavior is mediated by this perceived relevance rather than by algorithmic performance alone. This reveals a persistent and under-explored gap between algorithmic optimization and end-user experiences. From a technical perspective, automated metrics offer a standardized means of evaluating and comparing algorithmic advances, enabling researchers and developers to incrementally push the boundaries of algorithmic capabilities. However, improvements in such metrics do not reliably translate into corresponding gains in user preference or acceptance. As observed in Papers III and IV, although large language models outperform simpler approaches on commonly used benchmarking metrics, they can be outperformed in user-facing evaluations by those simpler methods that exhibit substantially lower automated evaluation scores. These findings do not imply that simpler methods are universally preferable. Rather, they suggest that the critical question is not *which* model performs best but *how* and *when* different approaches should be deployed. Viewed within the broader context of contextual communication supported by Paper V, effective system design requires consideration of available resources, intended goals (results), rectitude constraints and the needs of the target audience. In the advertising domain, this implies that advertisers should not merely chase the technological strategies of predominant large platforms, but instead assess their own resource availability, understand user preferences and expectations, and adopt transparent and privacy-aware practices that align with regulatory requirements.

The role of transparency, explored in greater depth in Paper IV, further underscores the importance of situating technical solutions within practical application contexts. While transparency mechanisms can enhance user-perceived relevance by offering a solid reference for ad placement, they may also raise privacy concerns or trigger senses of algorithmic intrusiveness. Moreover, the findings indicate that transparency can interact with algorithmic performance and user perceptions in nontrivial ways. In particular, when transparency explanations are provided, large language models exhibit more stable user preference patterns. This observation should not be interpreted as a universal conclusion, but rather as an illustrative case that highlights the sensitivity of user perception to explanation design. Open questions remain regarding how different forms of transparency explanation, beyond that used in the paper, may influence user responses and trust.

Overall, these results suggest that evaluating contextual language processing solely through automated metrics is insufficient for user-facing applications. Beyond precision- and efficiency-oriented algorithmic development and optimization, the effectiveness of contextual language processing systems must

be examined through user-centered evaluation within practical contexts. Such evaluations are essential for understanding how contextual relevance is perceived, interpreted, and accepted by users, and for ensuring that algorithmic advances translate into meaningful real-world impact.

## 5.3 Limitation and Future Work

Paper I opens the possibility of applying reinforcement learning to the grammatical inference of deterministic, non-probabilistic finite-state transducers. Paper II extends this line of work by incorporating deep RL techniques and addressing probabilistic FSTs. While these contributions establish a foundation for adaptive transducer learning, they also reveal some limitations that motivate future research.

In Paper II, the learning problem is decomposed into two primary subproblems: constructing a reference transducer structure and subsequently approximating transition probabilities. The first stage abstracts the target probabilistic transducer as a collection of deterministic, non-probabilistic transducer instances, enabling the identification of a set of optimal transduction actions for the same state. This abstraction simplifies the learning objective for tabular Q-learning methods and provides a practical compromise between the difficulty of representing multiple optimal actions in tabular methods and the increased complexity faced by neural approaches when state and action spaces evolve dynamically through state splitting.

However, this formulation relies on specific assumptions about the learning environment (i.e., the target transducer and sample strings). In particular, it assumes a constrained interaction, in which alternative actions at a given state are explored episodically through repeated interactions. In the transducer learning setting, this is enforced by a constraint rule of the environment that restricts non-alternative transition paths, effectively requiring the learner to interact with a one-way trajectory through the deterministic state-action combinations. While this constraint facilitates stable learning, it also limits the generality of the approach. Future work could explore more flexible learning frameworks that relax such environmental constraints, enabling transducer learning directly from randomly sampled input-output string pairs without relying on structured interaction protocols.

Beyond methodological limitations in transducer learning, additional opportunities arise from the user-centered aspects examined in Papers III and IV. Incorporating additional models and gold standards to compare their impacts would be a promising direction for future work. In addition, the transparency mechanisms investigated in Paper IV represent only one form of explanation design. Given increasing regulatory attention to data protection and growing user awareness of their privacy rights, the development of effective and trustworthy transparency explanations is likely to become an increasingly important research focus. Although prior work has identified certain factors that influ-

ence explanation effectiveness from user perspectives, a unified framework for designing and evaluating them remains under-explored. Future research could therefore investigate alternative explanation formats, levels of granularity, and personalization strategies, with the goal of deriving more generalizable insights into how transparency influences perceived relevance and trust.

Finally, a broader and more ambitious direction for future work lies in integrating reinforcement learning-based transducer learning with real-world application scenarios, with inspirations from the contextual advertising case studies. For instance, personalized contextual dialogue systems could be trained using reinforcement learning to acquire context-sensitive language behaviors, while optimizing system responses through user-centered relevance preference feedback. Such extensions would further strengthen the connection between adaptive contextual language processing models and their practical deployment in dynamic, user-centered environments.

## 5.4 Conclusion

This thesis examines contextual language processing as a fundamental part of AI-driven systems from both the methodological and application perspectives. It shows that effective use of contexts requires more than increased model complexity or improved benchmark performance. Instead, contextual language processing is best understood as an interpretable, adaptive, and user-centered interaction process.

Methodologically, the thesis demonstrates that reinforcement learning provides an adaptive and interpretable way for learning finite-state string transducers through string samples. By formulating grammatical inference as a sequential decision-making problem, contextual dependencies can be learned adaptively while retaining structural interpretability. This approach offers a practical alternative to purely rule-based systems and opaque neural models.

From the application point of view, studies in contextual advertising reveal a consistent gap between system-level relevance and user-perceived relevance. Improvements in automated benchmarking metrics or model sophistication do not necessarily translate into better user experiences. Instead, user perception, transparency, and trust play a central role in determining the effectiveness of such contextual communication.

Overall, this work highlights that advancing contextual language processing requires a shift in perspective: from precision-driven modeling toward interpretable, adaptive, and human-aligned interaction. By grounding technical innovation in both formal structure and human perception, the thesis contributes to the development of AI systems that are not only effective but also understandable, trustworthy, and meaningful in real-world contexts.

# Bibliography


[1] Robert J. Aalberts, Alexander Nill, and Percy S. Poon. "Online Behavioral Targeting: What Does the Law Say?" In: *Journal of Current Issues & Research in Advertising* 37.2 (2016), pp. 95–112. URL: https://doi.org/10.1080/10641734.2016.1171177.

[2] Haisal Dauda Abubakar, Mahmood Umar, and Muhammad Abdullahi Bakale. "Sentiment classification: Review of text vectorization methods: Bag of words, Tf-Idf, Word2vec and Doc2vec". In: *SLU Journal of Science and Technology* 4.1 (2022), pp. 27–33. URL: https://slujst.com.ng/wp-content/uploads/2023/09/266_pp_27-33.pdf.

[3] Yasser Alhelaly, Gurpreet Dhillon, and Tiago Oliveira. "The Impact of Privacy Intrusiveness on Individuals' Responses and Engagement Toward Personalization in Online Interactive Advertising". In: *Journal of Interactive Advertising* 25.1 (2025), pp. 38–60. URL: https://doi.org/10.1080/15252019.2024.2440318.

[4] Felipe Almeida and Geraldo Xexéo. *Word Embeddings: A Survey*. 2023. arXiv: 1901.09069 [cs.CL]. URL: https://arxiv.org/abs/1901.09069.

[5] Athanasios Andreou et al. "Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations". In: *NDSS 2018 - Network and Distributed System Security Symposium*. San Diego, United States, Feb. 2018, pp. 1–15. URL: https://doi.org/10.14722/ndss.2018.23204.

[6] D.M. Anisuzzaman et al. "Fine-Tuning Large Language Models for Specialized Use Cases". In: *Mayo Clinic Proceedings: Digital Health* 3.1 (2025), p. 100184. URL: https://doi.org/10.1016/j.mcpdig.2024.11.005.

[7] Juan Carlos Augusto. "Contexts and Context-awareness Revisited from an Intelligent Environments Perspective". In: *Applied Artificial Intelligence* 36.1 (2022), p. 2008644. URL: https://doi.org/10.1080/08839514.2021.2008644.

[8] Tae Hyun Baek and Mariko Morimoto. “Stay Away From Me”. In: *Journal of Advertising* 41.1 (2012), pp. 59–76. URL: https://doi.org/10.2753/JOA0091-3367410105.

[9] Raphaël Bailly, François Denis, and Liva Ralaivola. “Grammatical inference as a principal component analysis problem”. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML ’09. Montreal, Quebec, Canada: Association for Computing Machinery, 2009, pp. 33–40. URL: https://doi.org/10.1145/1553374.1553379.

[10] Borja Balle, Ariadna Quattoni, and Xavier Carreras. “A Spectral Learning Algorithm for Finite State Transducers”. In: *Machine Learning and Knowledge Discovery in Databases*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 156–171. URL: https://doi.org/10.1007/978-3-642-23780-5_20.

[11] Natã M. Barbosa et al. “Who Am I? A Design Probe Exploring Real-Time Transparency about Online and Offline User Profiling Underlying Targeted Ads”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5.3 (Sept. 2021). URL: https://doi.org/10.1145/3478122.

[12] Sophie C. Boerman, Sanne Kruikemeier, and Frederik J. Zuiderveen Borgesius. “Online Behavioral Advertising: A Literature Review and Research Agenda”. In: *Journal of Advertising* 46.3 (2017), pp. 363–376. URL: https://doi.org/10.1080/00913367.2017.1339368.

[13] Florian Boudin, Hugo Mougard, and Damien Cram. “How Document Pre-processing affects Keyphrase Extraction Performance”. In: *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 121–128. URL: https://aclanthology.org/W16-3917/.

[14] Andrei Broder et al. “A semantic approach to contextual advertising”. In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’07. Amsterdam, The Netherlands: Association for Computing Machinery, 2007, pp. 559–566. URL: https://doi.org/10.1145/1277741.1277837.

[15] Timothy A Brown, Michael T Moore, et al. “Confirmatory factor analysis”. In: *Handbook of structural equation modeling* 361 (2012), p. 379.

[16] Jingwen Cai, Sara Leckner, and Johanna Björklund. “From precision to perception: Human-in-the-loop evaluation of keyword extraction for internet-scale contextual advertising”. In: *Information Systems* 138 (2026), p. 102665. URL: https://doi.org/10.1016/j.is.2025.102665.

[17] *California Consumer Privacy Act of 2018 (Assembly Bill No. 375)*. California Civil Code §§ 1798.100–1798.199. Enacted June 28, 2018; amended by California Privacy Rights Act (CPRA) in 2020. June 2018. URL: https://oag.ca.gov/privacy/ccpa.

[18] Erik Cambria and Bebo White. "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]". In: *IEEE Computational Intelligence Magazine* 9.2 (2014), pp. 48–57. URL: https://doi.org/10.1109/MCI.2014.2307227.

[19] Jingcun Cao, Xiaolin Li, and Lingling Zhang. "Is Relevancy Everything? A Deep-Learning Approach to Understand the Effect of Image-Text Congruence". In: *Management Science* 71.12 (2025), pp. 10579–10602. URL: https://doi.org/10.1287/mnsc.2022.01896.

[20] F. Casacuberta et al. "Some approaches to statistical and finite-state speech-to-speech translation". In: *Computer Speech Language* 18.1 (2004), pp. 25–47. URL: https://www.sciencedirect.com/science/article/pii/S0885230803000287.

[21] Francisco Casacuberta and Enrique Vidal. "Machine Translation with Inferred Stochastic Finite-State Transducers". In: *Computational Linguistics* 30.2 (June 2004), pp. 205–225. URL: https://doi.org/10.1162/089120104323093294.

[22] Deepayan Chakrabarti, Deepak Agarwal, and Vanja Josifovski. "Contextual advertising by combining relevance with click feedback". In: *Proceedings of the 17th International Conference on World Wide Web*. WWW '08. Beijing, China: Association for Computing Machinery, 2008, pp. 417–426. URL: https://doi.org/10.1145/1367497.1367554.

[23] Si Chen et al. "How does ad relevance affect consumers' attitudes toward personalized advertisements and social media platforms? The role of information co-ownership, vulnerability, and privacy cynicism". In: *Journal of Retailing and Consumer Services* 73 (2023), p. 103336. URL: https://doi.org/10.1016/j.jretconser.2023.103336.

[24] N. Chomsky. "Three models for the description of language". In: *IRE Transactions on Information Theory* 2.3 (1956), pp. 113–124. URL: https://doi.org/10.1109/TIT.1956.1056813.

[25] Alexander Clark. "Memory-Based Learning of morphology with stochastic transducers". In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. ACL '02. Philadelphia, Pennsylvania: Association for Computational Linguistics, 2002, pp. 513–520. URL: https://doi.org/10.3115/1073083.1073169.

[26] James M. Crick and David Crick. "Designing robust electronic surveys in marketing research". In: *Journal of Strategic Marketing* 33.5 (2025), pp. 559–572. URL: https://doi.org/10.1080/0965254X.2025.2468679.

[27] Micael Dahlén. "THE MEDIUM AS A CONTEXTUAL CUE: Effects of Creative Media Choice". In: *Journal of Advertising* 34.3 (2005), pp. 89–98. URL: https://doi.org/10.1080/00913367.2005.10639197.

[28] DeepSeek-AI et al. *DeepSeek-V3 Technical Report*. 2025. arXiv: 2412.19437 [cs.CL]. URL: https://arxiv.org/abs/2412.19437.

[29] Jacob Devlin et al. "BERT: Pre-training of deep bidirectional transformers for language understanding". In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*. Vol. 1. 2019, pp. 4171–4186. URL: https://doi.org/10.18653/V1/N19-1423.

[30] Anind K. Dey. "Understanding and Using Context". In: *Personal Ubiquitous Comput.* 5.1 (Jan. 2001), pp. 4–7. URL: https://doi.org/10.1007/s007790170019.

[31] Ned Dimitrov. *Drive Contextual Advertising Results With Keyword Rule Targeting*. 2022. URL: https://www.stackadapt.com/resources/blog/keyword-rule-targeting.

[32] Qingxiu Dong et al. "A Survey on In-context Learning". In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 1107–1128. URL: https://doi.org/10.18653/v1/2024.emnlp-main.64.

[33] P. Dupont, F. Denis, and Y. Esposito. "Links between probabilistic automata and hidden Markov models: probability distributions, learning models and induction algorithms". In: *Pattern Recognition* 38.9 (2005), pp. 1349–1371. URL: https://doi.org/10.1016/j.patcog.2004.03.020.

[34] Jason Eisner. "Parameter Estimation for Probabilistic Finite-State Transducers". In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, July 2002, pp. 1–8. URL: https://aclanthology.org/P02-1001/.

[35] Motahhare Eslami et al. "Communicating Algorithmic Process in Online Behavioral Advertising". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–13. URL: https://doi.org/10.1145/3173574.3174006.

[36] Wenfang Fan, Bingjia Shao, and Yong Zhang. "The more congruent, the better? The role of recommendation content congruence on consumers' click-through intention in in-feed advertising". In: *Journal of Retailing and Consumer Services* 87 (2025), p. 104463. URL: https://doi.org/10.1016/j.jretconser.2025.104463.

[37] FasterCapital. *Contextual advertising and keyword matching: How to use keyword matching to improve your contextual advertising campaigns.* 2025. URL: https://fastercapital.com/content/Contextual-advertising-and-keyword-matching--How-to-use-keyword-matching-to-improve-your-contextual-advertising-campaigns.html.

[38] Nazanin Firoozeh et al. "Keyword extraction: Issues and methods". In: *Natural Language Engineering* 26.3 (2020), pp. 259–291. URL: https://doi.org/10.1017/S1351324919000457.

[39] W. Tecumseh Fitch and Angela D. Friederici. "Artificial grammar learning meets formal language theory: an overview". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1598 (2012), pp. 1933–1955. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2012.0103.

[40] Todd Michael Franke, Timothy Ho, and Christina A. Christie. "The Chi-Square Test: Often Used and More Often Misinterpreted". In: *American Journal of Evaluation* 33.3 (2012), pp. 448–458. URL: https://doi.org/10.1177/1098214011426594.

[41] Keith Frankish and Jonathan St. B. T. Evans. "1The duality of mind: An historical perspective". In: *In two minds: Dual processes and beyond.* Oxford University Press, Jan. 2009. URL: https://doi.org/10.1093/acprof:oso/9780199230167.003.0001.

[42] Marian Friestad and Peter Wright. "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts". In: *Journal of Consumer Research* 21.1 (1994), pp. 1–31. URL: http://www.jstor.org/stable/2489738.

[43] King-Sun Fu and Taylor L. Booth. "Grammatical Inference: Introduction and Survey - Part I". In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-5.1 (1975), pp. 95–111. URL: https://doi.org/10.1109/TSMC.1975.5409159.

[44] Mark Gales and Steve Young. "The Application of Hidden Markov Models in Speech Recognition". In: *Foundations and Trends in Signal Processing* 1.3 (2008), pp. 195–304. URL: http://dx.doi.org/10.1561/2000000004.

[45] Claas Christian Germelmann et al. "Congruence and Incongruence in Thematic Advertisement–Medium Combinations: Role of Awareness, Fluency, and Persuasion Knowledge". In: *Journal of Advertising* 49.2 (2020), pp. 141–164. URL: https://doi.org/10.1080/00913367.2020.1745110.

[46] Zhabiz Gharibshah and Xingquan Zhu. "User Response Prediction in Online Advertising". In: *ACM Comput. Surv.* 54.3 (May 2021). URL: https://doi.org/10.1145/3446662.

[47] Aaron Grattafiori et al. *The Llama 3 Herd of Models*. 2024. arXiv: 2407.21783 [cs.AI]. URL: https://arxiv.org/abs/2407.21783.

[48] Alex Graves. *Sequence Transduction with Recurrent Neural Networks*. 2012. arXiv: 1211.3711 [cs.NE]. URL: https://arxiv.org/abs/1211.3711.

[49] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. *Speech Recognition with Deep Recurrent Neural Networks*. 2013. arXiv: 1303.5778 [cs.NE]. URL: https://arxiv.org/abs/1303.5778.

[50] Alex Graves et al. "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks". In: *Proceedings of the 23rd International Conference on Machine Learning*. ICML '06. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 2006, pp. 369–376. URL: https://doi.org/10.1145/1143844.1143891.

[51] Alicja Grochowska and Andrzej Falkowski. "External and Internal Context in the Coherence of Advertisement: An Empirical Study on the Affect to the Ad and Brand". In: *Advances in Advertising Research VIII: Challenges in an Age of Dis-Engagement*. Wiesbaden: Springer Fachmedien Wiesbaden, 2017, pp. 97–107. URL: https://doi.org/10.1007/978-3-658-18731-6_8.

[52] Maarten Grootendorst. *KeyBERT: Minimal keyword extraction with BERT. v0.3.0 [software], Zenodo*. Version v0.3.0. 2020. URL: https://doi.org/10.5281/zenodo.4461265.

[53] Zhengrong Gu, Garrett Johnson, and Shunto Kobayashi. "Can Privacy Technologies Replace Cookies? Ad Revenue in a Field Experiment". In: (2025). URL: http://dx.doi.org/10.2139/ssrn.5319372.

[54] Tuomas Haarnoja et al. "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor". In: *Proceedings of the 35th International Conference on Machine Learning*. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 1861–1870. URL: https://proceedings.mlr.press/v80/haarnoja18b.html.

[55] E. Häglund and J. Björklund. "AI-Driven Contextual Advertising: Toward Relevant Messaging Without Personal Data". In: *Journal of Current Issues & Research in Advertising* 45.3 (2024), pp. 301–319. URL: https://doi.org/10.1080/10641734.2024.2334939.

[56] Anna Hätty, Simon Tannert, and Ulrich Heid. "Creating a gold standard corpus for terminological annotation from online forum data". In: *Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017)*. Montpellier, France: Association for Computational Linguistics, Sept. 2017. URL: https://aclanthology.org/W17-7002/.

[57] Yanzhang He et al. “Streaming End-to-end Speech Recognition for Mobile Devices”. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019, pp. 6381–6385. URL: https://doi.org/10.1109/ICASSP.2019.8682336.

[58] Susan E. Heckler and Terry L. Childers. “The Role of Expectancy and Relevancy in Memory for Verbal and Visual Information: What is Incongruency?” In: *Journal of Consumer Research* 18.4 (1992), pp. 475–492. URL: http://www.jstor.org/stable/2489260.

[59] Julia Hirschberg and Christopher D. Manning. “Advances in natural language processing”. In: *Science* 349.6245 (2015), pp. 261–266. URL: https://www.science.org/doi/abs/10.1126/science.aaa8685.

[60] Yue Huang et al. “Position: TRUSTLLM: trustworthiness in large language models”. In: *Proceedings of the 41st International Conference on Machine Learning*. ICML’24. Vienna, Austria: JMLR.org, 2024.

[61] Mans Hulden et al. “PyFoma: a Python finite-state compiler module”. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 258–265. URL: https://doi.org/10.18653/v1/2024.acl-demos.24.

[62] Anette Hulth. “Improved automatic keyword extraction given more linguistic knowledge”. In: *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. EMNLP ’03. USA: Association for Computational Linguistics, 2003, pp. 216–223. URL: https://doi.org/10.3115/1119355.1119383.

[63] Jitsuko Igarashi. “Contextual Language Processing Model”. In: *Proceedings of ’84 Matsuyama Workshop on Formal Grammar*. Waseda University. 1985, pp. 23–29.

[64] Janet Ilieva, Steve Baron, and Nigel M. Healey. “Online Surveys in Marketing Research”. In: *International Journal of Market Research* 44.3 (2002), pp. 1–14. URL: https://doi.org/10.1177/147078530204400303.

[65] Ganesh Iyer, David Soberman, and J. Miguel Villas-Boas. “The Targeting of Advertising”. In: *Marketing Science* 24.3 (2005), pp. 461–476. URL: https://doi.org/10.1287/mksc.1050.0117.

[66] Gerhard Jäger and James Rogers. “Formal language theory: refining the Chomsky hierarchy”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1598 (2012), pp. 1956–1970. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2012.0077.

[67] Navdeep Jaitly et al. *A Neural Transducer*. 2016. arXiv: 1511.04868 [cs.LG]. URL: https://arxiv.org/abs/1511.04868.

[68] Jim Jansen and Simone Schuster. "Bidding on the buying funnel for sponsored search and keyword advertising". In: *Journal of Electronic Commerce Research* 12 (Jan. 2011), pp. 1–18. URL: https://api.semanticscholar.org/CorpusID:328123.

[69] Albert Q. Jiang et al. *Mistral 7B*. 2023. arXiv: 2310.06825 [cs.CL]. URL: https://arxiv.org/abs/2310.06825.

[70] A-Reum Jung. "The influence of perceived ad relevance on social media advertising: An empirical examination of a mediating role of privacy concern". In: *Computers in Human Behavior* 70 (2017), pp. 303–309. URL: https://doi.org/10.1016/j.chb.2017.01.008.

[71] Maurits Kaptein and Judy Robertson. "Rethinking statistical analysis methods for CHI". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 1105–1114. URL: https://doi.org/10.1145/2207676.2208557.

[72] Sabrina Karwatzki et al. "Beyond the Personalization–Privacy Paradox: Privacy Valuation, Transparency Features, and Service Personalization". In: *Journal of Management Information Systems* 34.2 (2017), pp. 369–400. URL: https://doi.org/10.1080/07421222.2017.1334467.

[73] André Kempe. "Finite state transducers approximating Hidden Markov Models". In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. ACL '98/EACL '98. Madrid, Spain: Association for Computational Linguistics, 1997, pp. 460–467. URL: https://doi.org/10.3115/976909.979676.

[74] Maurice G. Kendall. *Rank Correlation Methods*. First. Griffin, 1949.

[75] Dhawal Khem, Shailesh D. Panchal, and C. B. Bhatt. "An Overview of Context Capturing Techniques in NLP". In: *International Journal on Recent and Innovation Trends in Computing and Communication* (2023). URL: https://api.semanticscholar.org/CorpusID:258755049.

[76] Tami Kim, Kate Barasz, and Leslie K John. "Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness". In: *Journal of Consumer Research* 45.5 (May 2018), pp. 906–932. URL: https://doi.org/10.1093/jcr/ucy039.

[77] Anastasia Kononova and Shupei Yuan. "Double-Dipping Effect? How Combining YouTube Environmental PSAs With Thematically Congruent Advertisements in Different Formats Affects Memory and Attitudes". In: *Journal of Interactive Advertising* 15.1 (2015), pp. 2–15. URL: https://doi.org/10.1080/15252019.2015.1009524.

[78] Anastasia Kononova et al. “Click, click, ad: the proportion of relevant (vs. irrelevant) ads matters when advertising within paginated online content”. In: *International Journal of Advertising* 39.7 (2020), pp. 1031–1058. URL: https://doi.org/10.1080/02650487.2020.1732114.

[79] Solomon Kullback. “Kullback-leibler divergence”. In: *Tech. Rep.* (1951).

[80] Shankar Kumar, Yonggang Deng, and William Byrne. “A weighted finite state transducer translation template model for statistical machine translation”. In: *Natural Language Engineering* 12.1 (2006), pp. 35–75. URL: https://doi.org/10.1017/S1351324905003815.

[81] Kim Tae Kyun. “T test as a parametric statistic”. In: *kja* 68.6 (2015), pp. 540–546. URL: https://doi.org/10.4097/kjae.2015.68.6.540.

[82] Kuang-chih Lee et al. “Estimating conversion rate in display advertising from past erformance data”. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’12. Beijing, China: Association for Computing Machinery, 2012, pp. 768–776. URL: https://doi.org/10.1145/2339530.2339651.

[83] Randall Lewis, Justin M. Rao, and David H. Reiley. “Measuring the Effects of Advertising: The Digital Frontier”. In: *Economic Analysis of the Digital Economy*. NBER Chapters. National Bureau of Economic Research, Inc, Dec. 2015, pp. 191–218. URL: https://ideas.repec.org/h/nbr/nberch/12991.html.

[84] Huihan Li et al. “Ditch the Gold Standard: Re-evaluating Conversational Question Answering”. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 8074–8085. URL: https://doi.org/10.18653/v1/2022.acl-long.555.

[85] Huiran Li and Yanwu Yang. “Keyword targeting optimization in sponsored search advertising: Combining selection and matching”. In: *Electronic Commerce Research and Applications* 56 (2022), p. 101209. URL: https://doi.org/10.1016/j.elerap.2022.101209.

[86] Q. Vera Liao and Jennifer Wortman Vaughan. *AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap*. 2023. arXiv: 2306.01941 [cs.HC]. URL: https://arxiv.org/abs/2306.01941.

[87] Elizabeth D Liddy. “Natural language processing”. In: (2001).

[88] Chen Ling et al. “Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey”. In: *ACM Comput. Surv.* 58.3 (Oct. 2025). URL: https://doi.org/10.1145/3764579.

[89] Pengqi Liu, Javad Azimi, and Ruofei Zhang. “Automatic keywords generation for contextual advertising”. In: *Proceedings of the 23rd International Conference on World Wide Web*. WWW ’14 Companion. Seoul, Korea: Association for Computing Machinery, 2014, pp. 345–346. URL: https://doi.org/10.1145/2567948.2577361.

[90] Yixin Liu et al. “Revisiting the Gold Standard: Grounding Summarization Evaluation with Robust Human Evaluation”. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 4140–4170. URL: https://doi.org/10.18653/v1/2023.acl-long.228.

[91] Xianghua Lu, Xia Zhao, and Ling Xue. “Is Combining Contextual and Behavioral Targeting Strategies Effective in Online Advertising?” In: *ACM Trans. Manage. Inf. Syst.* 7.1 (Feb. 2016). URL: https://doi.org/10.1145/2883816.

[92] Virginie Maille and Nathalie Fleck. “Perceived Congruence and Incongruence: Toward a Clarification of the Concept, its Formation and Measure”. In: *Recherche et Applications en Marketing* 26.2 (2011), pp. 77–113. URL: https://doi.org/10.1177/205157071102600204.

[93] Nacef Ben Mansour, Hamed Rahimi, and Motasem Alrahabi. “How Well Do Large Language Models Extract Keywords? A Systematic Evaluation on Scientific Corpora”. In: *Proceedings of the 1st Workshop on AI and Scientific Discovery: Directions and Opportunities*. Albuquerque, New Mexico, USA: Association for Computational Linguistics, May 2025, pp. 13–21. URL: https://doi.org/10.18653/v1/2025.aisd-main.2.

[94] Jiaxin Mao et al. “When does Relevance Mean Usefulness and User Satisfaction in Web Search?” In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’16. Pisa, Italy: Association for Computing Machinery, 2016, pp. 463–472. URL: https://doi.org/10.1145/2911451.2911507.

[95] Thierry Maout. *Google Chrome is keeping third-party cookies after all: What does it mean?* Online. Apr. 2025. URL: https://www.didomi.io/blog/google-chrome-third-party-cookies-april-2025.

[96] Rui Meng et al. “Deep Keyphrase Generation”. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, July 2017, pp. 582–592. URL: https://doi.org/10.18653/v1/P17-1054.

[97] Yajie Miao, Mohammad Gowayyed, and Florian Metze. *EESEN: End-to-End Speech Recognition using Deep RNN Models and WFST-based Decoding*. 2015. arXiv: 1507.08240 [cs.CL]. URL: https://arxiv.org/abs/1507.08240.

[98] Tomás Mikolov et al. "Efficient Estimation of Word Representations in Vector Space". In: *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. 2013. URL: http://arxiv.org/abs/1301.3781.

[99] Mehryar Mohri. "Finite-State Transducers in Language and Speech Processing". In: *Computational Linguistics* 23.2 (1997), pp. 269–311. URL: https://aclanthology.org/J97-2003/.

[100] Mehryar Mohri, Fernando Pereira, and Michael Riley. "Weighted finite-state transducers in speech recognition". In: *Computer Speech Language* 16.1 (2002), pp. 69–88. URL: https://www.sciencedirect.com/science/article/pii/S0885230801901846.

[101] Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. "Natural language processing: an introduction". In: *Journal of the American Medical Informatics Association* 18.5 (Sept. 2011), pp. 544–551. URL: https://doi.org/10.1136/amiajnl-2011-000464.

[102] Navakanth Reddy Naredla and Festus Fatai Adedoyin. "Detection of Hyperpartisan news articles using natural language processing techniques". In: *International Journal of Information Management Data Insights* 2.1 (2022).

[103] Graham Neubig. "Neural Machine Translation and Sequence-to-sequence Models: A Tutorial". In: (Mar. 2017). URL: https://doi.org/10.48550/arXiv.1703.01619.

[104] Juri Opitz, Shira Wein, and Nathan Schneider. "Natural Language Processing RELIES on Linguistics". In: *Computational Linguistics* 51.3 (Sept. 2025), pp. 1009–1032. URL: https://doi.org/10.1162/coli_a_00560.

[105] Takeru Otani et al. "Speaker-dependent Continuous Speech Recognition for Individuals with Cerebral Palsy Using Weighted Finite-State Transducer and Text-to-Speech Synthesis". In: *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*. ASSETS '25. Association for Computing Machinery, 2025. URL: https://doi.org/10.1145/3663547.3759757.

[106] Ayse Bengi Ozcelik and Kaan Varnali. "Effectiveness of online behavioral targeting: A psychological perspective". In: *Electronic Commerce Research and Applications* 33 (2019), p. 100819. URL: https://doi.org/10.1016/j.elerap.2018.11.006.

[107] Dae-Yun Park. “Enhancing Credibility in Programmatic Advertising: A Conceptual Model of Transparency, Optimality, and Safety Surveillance”. In: *Journalism & Mass Communication Quarterly* (2025). URL: https://doi.org/10.1177/10776990251328625.

[108] Jiaxin Pei et al. “POTATO: The Portable Text Annotation Tool”. In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Abu Dhabi, UAE: Association for Computational Linguistics, 2022, pp. 327–337. URL: https://aclanthology.org/2022.emnlp-demos.33.

[109] Patrick De Pelsmacker, Maggie Geuens, and Pascal Anckaert. “Media Context and Advertising Effectiveness: The Role of Context Appreciation and Context/Ad Similarity”. In: *Journal of Advertising* 31.2 (2002), pp. 49–61. URL: http://www.jstor.org/stable/4189214.

[110] Do Viet Phuong and Tu Minh Phuong. “A keyword-topic model for contextual advertising”. In: *Proceedings of the 3rd Symposium on Information and Communication Technology*. SoICT ’12. Ha Long, Vietnam: Association for Computing Machinery, 2012, pp. 63–70. URL: https://doi.org/10.1145/2350716.2350728.

[111] Daniel Povey et al. “Generating exact lattices in the WFST framework”. In: 2012, pp. 4213–4216. URL: https://doi.org/10.1109/ICASSP.2012.6288848.

[112] Wisam A. Qader, Musa M. Ameen, and Bilal I. Ahmed. “An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges”. In: *2019 International Engineering Conference (IEC)*. 2019, pp. 200–204. URL: https://doi.org/10.1109/IEC47844.2019.8950616.

[113] Kanishka Rao, Haşim Sak, and Rohit Prabhavalkar. “Exploring architectures, data and units for streaming end-to-end speech recognition with RNN-transducer”. In: *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. 2017, pp. 193–199. URL: https://doi.org/10.1109/ASRU.2017.8268935.

[114] Pushpendre Rastogi, Ryan Cotterell, and Jason Eisner. “Weighting Finite-State Transductions With Neural Context”. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics, June 2016, pp. 623–633. URL: https://aclanthology.org/N16-1076/.

[115] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. Official Journal of the European Union, L 119, pp. 1–88. Entered into application on 25 May 2018. May 2016.

[116] Matthew Richardson, Ewa Dominowska, and Robert Ragno. "Predicting clicks: estimating the click-through rate for new ads". In: *Proceedings of the 16th International Conference on World Wide Web*. WWW '07. Banff, Alberta, Canada: Association for Computing Machinery, 2007, pp. 521–530. URL: https://doi.org/10.1145/1242572.1242643.

[117] Amin Sayedi. "Real-Time Bidding in Online Display Advertising". In: *Marketing Science* 37.4 (2018), pp. 553–568. URL: https://doi.org/10.1287/mksc.2017.1083.

[118] Dietram A. Scheufele and Tewksbury David. "Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models". In: *Journal of Communication* 57.1 (2007), pp. 9–20. URL: https://doi.org/10.1111/j.0021-9916.2007.00326.x.

[119] Viktor Schlegel et al. "A Framework for Evaluation of Machine Reading Comprehension Gold Standards". eng. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 5359–5369. URL: https://aclanthology.org/2020.lrec-1.660/.

[120] Sigal Segev, Weirui Wang, and Juliana Fernandes. "The effects of ad–context congruency on responses to advertising in blogs". In: *International Journal of Advertising* 33.1 (2014), pp. 17–36. URL: https://doi.org/10.2501/IJA-33-1-017-036.

[121] Noah A. Smith. "Contextual word representations: putting words into computers". In: *Commun. ACM* 63.6 (May 2020), pp. 66–74. URL: https://doi.org/10.1145/3347145.

[122] Karen Sparck Jones. "A statistical interpretation of term specificity and its application in retrieval". In: *Document Retrieval Systems*. GBR: Taylor Graham Publishing, 1988, pp. 132–142.

[123] Horst Stip. "How Context Can Make Advertising More Effective". In: *Journal of Advertising Research* 58.2 (2018), pp. 138–145. URL: https://doi.org/10.2501/JAR-2018-022.

[124] Ching Y. Suen. "n-Gram Statistics for Natural Language Understanding and Text Processing". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 1.2 (Feb. 1979), pp. 164–172. URL: https://doi.org/10.1109/TPAMI.1979.4766902.

[125] Achyut Telang et al. "The Effects of Context Congruence On Ad Persuasiveness in e-Magazines". In: *Journal of Advertising Research* 61.3 (2021), pp. 303–317. URL: https://doi.org/10.2501/JAR-2021-002.

[126] Hugo Touvron et al. "Llama 2: Open Foundation and Fine-Tuned Chat Models". In: *arXiv preprint* (2023). URL: https://doi.org/10.48550/arXiv.2307.09288.

[127] Debra Trampe et al. “Beauty as a tool: The effect of model attractiveness, product relevance, and elaboration likelihood on advertising effectiveness”. In: *Psychology & Marketing* 27.12 (2010), pp. 1101–1121. URL: https://doi.org/10.1002/mar.20375.

[128] Junichi Tsujii. “Natural Language Processing and Computational Linguistics”. In: *Computational Linguistics* 47.4 (Dec. 2021), pp. 707–727. URL: https://doi.org/10.1162/coli_a_00420.

[129] Nicholas Tyacke and Rohan Higgins. “Searching for trouble – keyword advertising and trade mark infringement”. In: *Computer Law Security Report* 20.6 (2004), pp. 453–465. URL: https://doi.org/10.1016/S0267-3649(04)00090-1.

[130] Jodie B. Ullman and Peter M. Bentler. “Structural Equation Modeling”. In: *Handbook of Psychology, Second Edition*. John Wiley Sons, Ltd, 2012. Chap. 23. URL: https://doi.org/10.1002/9781118133880.hop202023.

[131] Kaan Varnali. “Online behavioral advertising: An integrative review”. In: *Journal of Marketing Communications* 27.1 (2021), pp. 93–114. URL: https://doi.org/10.1080/13527266.2019.1630664.

[132] Ashish Vaswani et al. “Attention is All you Need”. In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[133] Enrique Vidal and Francisco Casacuberta. “Learning Finite-State Models for Machine Translation”. In: *Grammatical Inference: Algorithms and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 3–15. URL: https://doi.org/10.1007/978-3-540-30195-0_2.

[134] Christopher JCH Watkins and Peter Dayan. “Q-learning”. In: *Machine learning* 8.3 (1992), pp. 279–292. URL: https://doi.org/10.1007/BF00992698.

[135] Jason Wei et al. “Chain-of-thought prompting elicits reasoning in large language models”. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS ’22. New Orleans, LA, USA: Curran Associates Inc., 2022. URL: https://dl.acm.org/doi/10.5555/3600270.3602070.

[136] Lars Wissler et al. “The Gold Standard in Corpus Annotation.” In: *IEEE GSC* 21 (2014). URL: https://doi.org/10.13140/2.1.4316.3523.

[137] Zongda Wu et al. “Leveraging Wikipedia concept and category information to enhance contextual advertising”. In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. CIKM ’11. Glasgow, Scotland, UK: Association for Computing Machinery, 2011, pp. 2105–2108. URL: https://doi.org/10.1145/2063576.2063901.

[138] Zongda Wu et al. “Position-wise contextual advertising: Placing relevant ads at appropriate positions of a web page”. In: *Neurocomputing* 120 (2013). Image Feature Detection and Description, pp. 524–535. URL: https://doi.org/10.1016/j.neucom.2013.04.018.

[139] Jun Yan et al. “How much can behavioral targeting help online advertising?” In: *Proceedings of the 18th International Conference on World Wide Web*. WWW ’09. Madrid, Spain: Association for Computing Machinery, 2009, pp. 261–270. URL: https://doi.org/10.1145/1526709.1526745.

[140] Jing Yang and Mengtian Jiang. “Demystifying congruence effects in Instagram in-feed native ads: the role of media-based and self-based congruence”. In: *Journal of Research in Interactive Marketing* 15.4 (June 2021), pp. 685–708. URL: https://doi.org/10.1108/JRIM-03-2020-0048.

[141] Idil Yaveroglu and Naveen Donthu. “Advertising Repetition and Placement Issues in On-Line Environments”. In: *Journal of Advertising* 37.2 (2008), pp. 31–44. URL: https://doi.org/10.2753/JOA0091-3367370203.

[142] Kaifu Zhang and Zsolt Katona. “Contextual Advertising”. In: *Marketing Science* 31.6 (2012), pp. 980–994. URL: https://doi.org/10.1287/mksc.1120.0740.

[143] Qian Zhang et al. “Transformer Transducer: A Streamable Speech Recognition Model with Transformer Encoders and RNN-T Loss”. In: *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 7829–7833. URL: https://doi.org/10.1109/ICASSP40776.2020.9053896.

[144] Haiyan Zhao et al. “Explainability for Large Language Models: A Survey”. In: *ACM Trans. Intell. Syst. Technol.* 15.2 (Feb. 2024). URL: https://doi.org/10.1145/3639372.