Supporting Creativity with Conversational Agents

The Influence of Perceived Age in AI-generated Voices on Creative Expression

Maria Sylvia Alaoja
Guðfinna Kristinsdóttir

Aalborg University
Department of Sustainability and Planning

This report was created using an AAU report template that Jesper Kjær Nielsen created, available on the Overleaf website. Additional packages were added to facilitate the IEEE citation system and Bibliography.

Department of Sustainability and Planning
Aalborg University
http://www.aau.dk

Title:
Supporting Creativity with Conversational Agents
The Influence of Perceived Age in AI-generated Voices on Creative Expression

Education:
Techno-Anthropology

Semester:
Spring 2025

Group number:
15

Authors:
Maria Sylvia Alaoja
Guðfinna Kristinsdóttir

Supervisors:
Christian Gradhandt Nøhr
Kristina Tornbjerg Eriksen

Copies: 1

Page Numbers: 55

Date of Completion:
May 28, 2025

Abstract:

This Master's thesis is based on a collaboration between Aalborg University's Department of Materials and Production, Robotics and Automation team, and LEGO®. By employing a custom-built voice-based interface and adapted creativity metrics, it explores how child-like versus adult-like AI-generated voices of conversational agents (CAs) affect verbal creativity in a play-based setting. Drawing on both qualitative and quantitative methodologies, we analyse the socio-technical conditions that shape creativity assessment. We critically examine the assumptions embedded in standardised creativity metrics through the lens of the Sociology of Scientific Knowledge (SSK), and propose a context-sensitive approach to measuring creativity. The findings from adult participants in this study point towards the notion that the perceived age of AI-generated voices in CAs can significantly influence the quantity and the quality of participants creative expression. These results establish a comparative baseline for a forthcoming study with children. Through this interdisciplinary approach, the project contributes to more responsible and reflective AI design for child users.

Contents

Preface **v**

Acknowledgements **vii**

1 Introduction **1**
1.1 Theoretical Grounding 3
1.2 Historical Context of AI 4
1.3 Surge of Conversational Agents 6
1.3.1 Children as Emerging Users of CAs 7
1.3.2 Perception of AI-generated Voices 8
1.4 Children Learn Creativity Through Play 9
1.5 Equality Supports Children's Agency 12
1.6 Structure Overview 14

2 Problem Analysis **15**
2.1 Creativity in the Context of HRI 15
2.2 Historical and Disciplinary Roots of Creativity Research 16
2.3 Challenges in Measuring Creativity 19
2.4 Problem Statement 23
2.5 Research Questions 23

3 Methodology **24**
3.1 Study Design and Participants 24
3.1.1 Participant profile 25
3.1.2 Study conditions 25
3.1.3 Game design 25
3.1.4 Turn sequence 26
3.1.5 Technical design of the interface 27
3.2 Study Setup 27
3.3 Data 28
3.3.1 Data collection 28
3.3.2 Data analysis 29

3.4 Ethical Considerations 29
3.5 Limitations 30

4 Results **31**
4.1 Data Description 31
4.1.1 Fluency and Elaboration 31
4.1.2 Distribution of Elaboration 32
4.2 Summary and Conclusions 35
4.3 Future Study Settings 36
4.3.1 Creativity Pre-Test 36
4.3.2 Rubric Scoring System of The Pre-Test 37

5 Discussion **40**
5.1 The Historical and Structural Biases of HRI Research 40
5.2 Missing Child Perspective 42
5.3 Connecting Metrics, Meaning and Machines 43
5.4 Limitations and reflections 44
5.5 Towards Future Approaches 46

Preface

This Techno-Anthropology Master's thesis explores the intersection of anthropological perspectives and Human-Robot Interaction (HRI), with a specific focus on Conversational Agents (CAs) and child-robot interaction (CRI). CAs are voice-based artificial intelligence (AI) systems that communicate through natural spoken language rather than text, offering a unique opportunity to investigate how humans engage with emerging technologies in social and developmental contexts.

The thesis is rooted in a collaborative research project between Aalborg University's Department of Materials and Production, specifically the Robotics and Automation group, and LEGO Systems A/S. This partnership was initiated during our final (10th) semester, building on internships completed the previous semester: one author at the Robotics and Automation Group, the other at LEGO Kids Engagement. These experiences sparked a shared interest in studying AI and robotics technologies in real-world settings, culminating in the project that we named "Learning Through Play with AI". The project is designed to examine how AI-generated child versus adult voices in CAs influence children's creative expression during play.

This collaboration bridges AAU's technical and quantitative expertise in designing HRI interfaces with The LEGO Group's child-centered, qualitative approach, which emphasizes the potential of emerging technologies to enrich play and nurture creativity. Our thesis is grounded in this interdisciplinary effort, combining technical innovation with a deep understanding of child development, creativity and play.

A guiding question from Aalborg University's Techno-Anthropology Master's program encapsulates the spirit of our work: "Do you want to take part in creating a responsible future with technology?" This question has shaped our academic journey and serves as a foundation for both this thesis and the broader research project. Through this work, we aim to contribute to knowledge that supports the responsible development and application of AI technologies in contexts involving children.

> "There is a lot of knowledge about human beings and how we interact with each other. But we are lacking knowledge of our interaction

> with technology—and how humans and technology mutually affect each other. As a Techno-Anthropologist, you will take part in creating this knowledge—thereby ensuring a future with responsible use of technology." (AAU, MSc programme in Techno-Anthropology)

The "Learning Through Play with AI" project specifically investigates the effects of AI-generated voice interfaces, comparing childlike and adult voices, on creative expression during gameplay involving a voice system. By including both adult and child participants, the study aims to reveal developmental differences in how these voices influence interaction and creativity.

This work is framed within the Learning through Play framework developed by the LEGO Foundation and UNICEF [1], which positions play as a vital context for learning, creativity, and development. The project strives to produce insights into how AI can best support children's creative expression while respecting their autonomy and diverse needs. A central objective is to determine whether conversational agents that use childlike AI voices encourage more imaginative and engaging play experiences.

This thesis follows a problem-based learning approach, emphasising real-world challenges and collaboration. By engaging directly with practical issues in HRI and AI-enhanced play, the research promotes active learning through problem-solving and critical reflection, aligning academic inquiry with technological and social contexts.

Aalborg University, May 28, 2025

Acknowledgements

This thesis would not have been possible without the support, guidance and collaboration of many individuals to whom we are deeply grateful.

We would like to express our sincere thanks to our Master's thesis supervisors from Aalborg University Department of Sustainability and Planning Professor Christian Gradhandt Nøhr, and Assistant Professor Kristina Tornbjerg Eriksen.

We wish to thank the dedicated team members from Aalborg University Robotics and Automation group involved in this project, for making the intersnhip and Master's thesis project possible, and for offering their knowledge and expertise on human-robot interaction, MLLMs, machine learning and natural language processing, generative conversational AI, as well as support and interest to make this project happen through active participation, help and guidance. We would especially like to thank Associate Professors Chen Li and Dimitris Chrysostomou, Intern and a Master's Student in Sungshin Women's University, Suzy Choi, Post-Doc David Andres Figueroa Salvador, and a student worker Anders Bloch Lauridsen, whose contributions, collaboration and shared commitment made this work a meaningful and enjoyable learning experience.

We also thank the LEGO Group for their enthusiasm regarding our research project and the immense knowledge and experience each of them shared openly with us about how to do research with children responsibly and ethically, and teaching us ways to make the experience as enjoyable as possible for everyone involved in the research. We would especially like to thank Henrik Daae, Jen Benz, Karina Kjærgaard and Jesper Soederberg, because without their expertise and the passion for children and play, this research project would not be happening.

And also additionally our special thanks are extended to the researchers and experts from related fields who generously offered their insights and expertise to guide us forward in out project. We are particularly thankful to Lecturer Lin Zou from De Montfort University, and Lecturer Lu Zhang from Beijing Union University, who helped clarify complex frameworks and whose consultation steered us confidently to right direction in creativity measurements, as well as to mathematician, Associate Professor Oliver W. Gnilke from Aalborg University who answered questions and provided guidance about mathematical aaspects of data analysis.

We'd like to thank our friends and family that supported us throughout the education. Guðfinna Kristinsdóttir would especially like to mention her husband, Hrannar Örn Karlsson, her daughter Elísabet Lillith Hrannarsdóttir, and her own mother Ingunn Sveinsdóttir.

Finally, we would like to thank each other for a smooth and inspiring collaboration, not limited to this particular semester, but also to our work together in previous semester projects: During our 7th with our work on predictive AI in police work in Denmark and Iceland, and 8th semester work on online horse riding and horse enthusiasts gaming communities built around Red Dead Online. Our shared enthusiasm for these topics continuously fueled our motivation, and the way our individual strengths complemented one another allowed us to contribute meaningfully to these projects. Working together in such a balanced and supportive way has been both rewarding and energising.

Chapter 1

Introduction

> Studying anthropology is like embarking on a journey which turns out to be much longer than you had initially planned, possibly because the plans were somewhat open-ended to begin with, and the terrain turned out to be bumpier and more diverse than the map suggested
> - Thomas Hylland Eriksen [2]

The Joint Research Centre (JRC) of the European Commission has identified creativity as a “transversal skill” - one of nine key competencies individuals should develop beyond foundational literacy and numeracy [3]. Importantly, creativity is not a fixed trait but a flexible skill that can be nurtured from early childhood onward [4]. This recognition highlights the growing emphasis on creativity as a fundamental component of lifelong learning. For understanding the evolving landscape of creativity research within technological disciplines is essential to situate this thesis in its broader academic context. A literature search conducted through the Scopus database reveals clear trends in how creativity and AI intersect across Engineering and Computer Science.

As illustrated in Figure 4.4, publications combining the keywords "creativity" and "AI" have grown steadily since the early 2000s, with a marked acceleration starting around 2017. This surge reflects the increasing attention paid to computational creativity, as well as AI’s expanding role in enabling, augmenting, and reshaping creative processes across diverse applications.

When narrowing the focus to research that explicitly addresses children within this intersection, as shown in Figure 1.2, the publication volume is considerably smaller but exhibits a similar upward trajectory. Starting with very few studies in the early 2000s, the field has gradually gained momentum since 2010, with a notable increase after 2018. This pattern points to an emerging recognition of the unique opportunities and challenges AI presents in child-centered domains such as education, play, and developmental support.

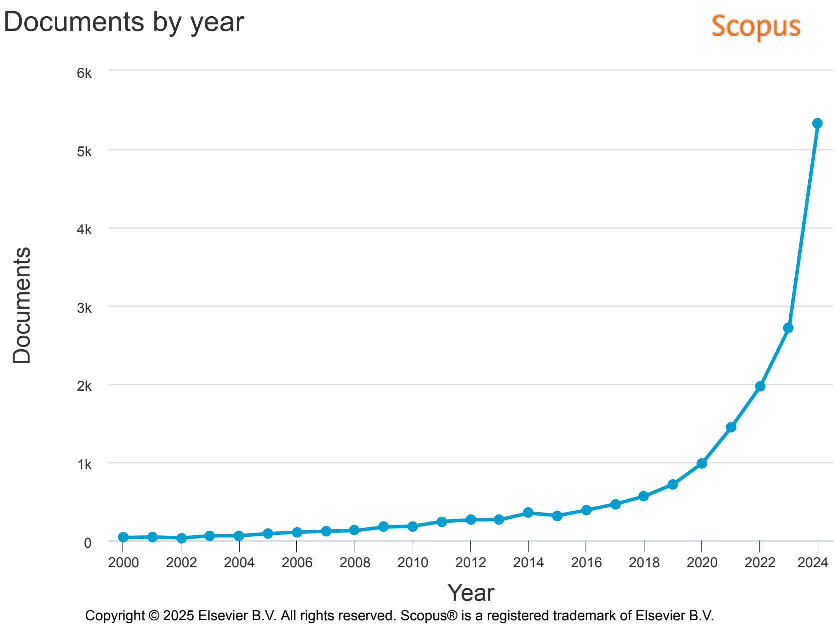

Figure 1.1: Articles that include both keywords "Creativity" and "AI" published between 2000-2025 in the fields of engineering and computer science

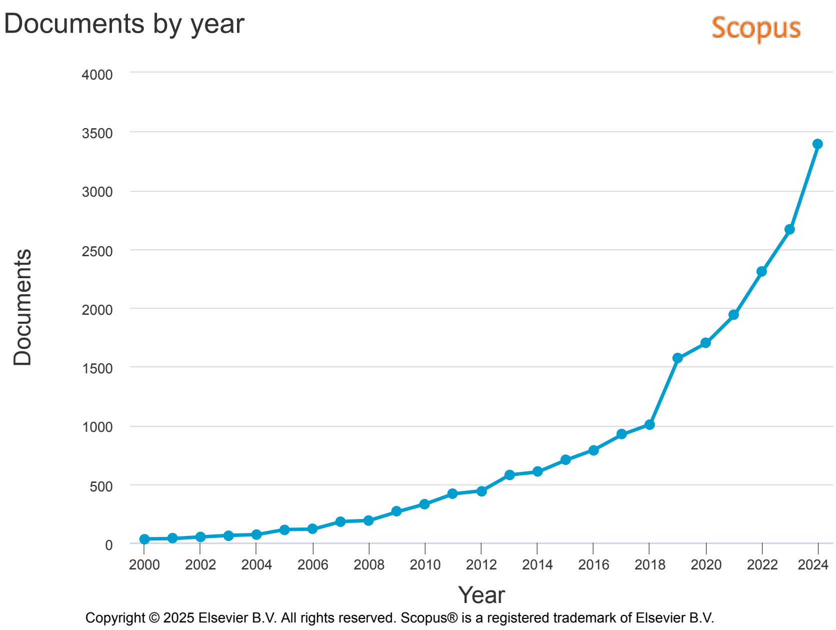

Figure 1.2: Articles that include all keywords "Creativity" and "AI" and "Children" published between 2000-2025 in the fields of engineering and computer science

Despite this growing interest, the intersection of AI, creativity, and children remains relatively underexplored, highlighting a significant gap in both research and practical design frameworks. This thesis positions itself within this dynamic and timely field, addressing critical questions about how AI-driven creative technologies can be effectively designed for and engaged with by children. By focusing on this emergent topic, the research aims to contribute novel insights that are highly relevant to current technological, educational, and societal developments.

In this chapter, we take an anthropological perspective by situating our research within its broader socio-cultural context. We begin with introducing the sociology of scientific knowlegde an overview of the historical development of artificial intelligence (AI) and examine how AI has been portrayed in Western media. Understanding this history is important because it has influenced the design of our study. We then connect these developments to current and emerging AI technologies, focusing specifically on their relevance to children and creative play.

1.1 Theoretical Grounding

Our study design and methodological choices were informed by interdisciplinary scientific research on children's creativity, cognitive development, social peer vs. adult interaction, and voice perception in HRI, alongside with a comprehensive review of creativity literature.

Beyond this empirical foundation, the study draws on broader epistemological insights from the *sociology of scientific knowledge* (SSK). SSK builds upon the foundational work of K. Mannheim (1936) in *the sociology of knowledge*. Initially, the sociology of knowledge treated scientific and mathematical knowledge as distinct from other forms of human understanding, often assuming that natural sciences produce objective, universally valid truths, unaffected by social, cultural, or historical contexts. While Mannheim acknowledged that many forms of knowledge—such as ideologies or belief systems—are socially situated, scientific knowledge was largely viewed as following a neutral trajectory grounded in logic and natural laws. [5]

As the field progressed, scholars critically examined this dichotomy. Contemporary SSK argues that scientific knowledge is also socially constructed and embedded within specific institutional, cultural, and historical contexts. These include influences such as research funding structures, disciplinary norms, methodological preferences, and broader sociopolitical environments shaping scientific inquiry. SSK does not deny the value of scientific methods but emphasizes that the production and interpretation of scientific facts are shaped by social processes and epistemic negotiations within scientific communities. [6]

Despite these insights, sociological analyses of the natural sciences often encounter resistance. Many scientific domains view social dimensions of knowledge production as secondary or external to core practices. The prevailing ideal prioritizes methodological control and internal validity, particularly in experimental disciplines. While such control is necessary within laboratory confines, applying scientific knowledge in real-world settings requires attention to broader social and contextual factors influencing its meaning and impact. [7]

Scientific knowledge production has traditionally been dominated by quantitative and statistical methods. These approaches excel at identifying patterns within large datasets and generating findings generalizable across contexts. They have been crucial for producing reliable and replicable results, especially in the natural sciences. As S. Yearley, a British sociologist specializing in SSK, notes that many researchers assume that scientific facts represent objective and singular truths, such as the exact molecular structure of DNA, or the precise dating of a fossil [8, p. 246]. This perspective reflects a commitment to uncovering a definitive version of reality through empirical evidence, and often overlooking the complexities and multiple interpretations in knowledge production.

Through this evidence-based, multiperspective approach, the study ensures its experimental design aligns with current interdisciplinary understandings while remaining critically aware of the socio-cultural dimensions underlying both human development and technological design. Following SSK, concepts such as creativity are understood not as fixed or universal but as socially and historically situated, shaped by disciplinary traditions, institutional contexts, and broader cultural and technological influences. The study maintains a reflexive stance toward its methodological choices, an approach particularly important in AI systems that implicitly carry normative assumptions about cognition, behavior, and interaction.

This reflexivity lays the groundwork for understanding the historical development of artificial intelligence and its socio-technical implications, which the next section addresses.

1.2 Historical Context of AI

Understanding the historical trajectory of AI is essential to grasp both the technological milestones and also the cultural and societal frameworks that shape how AI is developed, perceived, and integrated into everyday life. This context highlights how historical development of AI is not only a technical journey but it is intertwined with human values, imagination which are factors that continue to influence current innovations such as CAs.

AI is one of the most rapidly evolving and enigmatic fields in modern science. The past decade, and particularly the last three years, there has been exponential growth in AI development [9], yet the origins of AI extend far deeper into history. The story of AI is not only one of technological innovation but also a rich narrative shaped by human imagination, filled with dreams, promises, and speculative visions [10]. Its philosophical roots can be traced back to mechanistic philosophy, where early thinkers envisioned intelligent machines as mirrors reflecting human nature and cognition [10].

For centuries, humans have been fascinated by automata, mechanical devices designed to mimic life. One of the earliest examples is the wooden pigeon created around 400 BCE by Archytas of Tarentum. This device, powered by compressed air and balanced on a pivoting bar, demonstrated early attempts to simulate living motion through mechanical principles [11]. Such inventions highlight a growing belief that natural phenomena could be understood, modeled, and perhaps replicated using mechanistic laws [12], laying foundational ideas that would later influence scientific instrumentation and robotics.

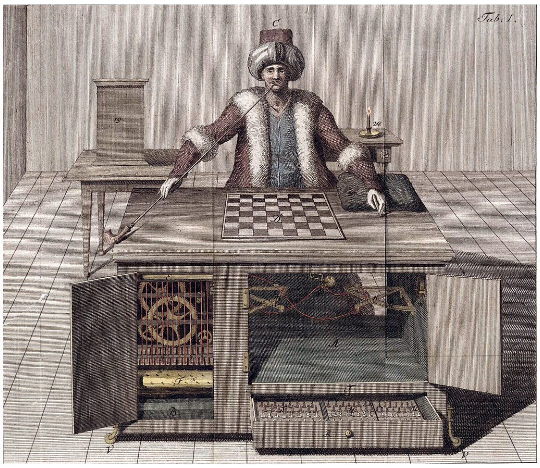

Figure 1.3: Drawing of The Turk by Joseph Racknitz, 1789

While early automata lacked true autonomy or intelligence, their cultural and technological influence was profound and far-reaching. A notable example is The Turk, an 18th-century chess-playing automaton constructed in 1770, that was wearing an "oriental" attire. Though it was ultimately revealed as a cleverly concealed illusion operated by a hidden human, The Turk captivated audiences for over eighty years, sparking imaginations about machine intelligence and automation. Its impact extended beyond entertainment, inspiring pioneers like C. Babbage. After losing to The Turk twice, Babbage was motivated to develop the Difference Engine, which is now recognized as one of the earliest mechanical computers [11].

This historical illusion of machine autonomy has a modern parallel in Amazon's 2005 launch of Mechanical Turk (AMT). AMT introduced a platform for crowdsourcing microtasks, which critics have characterized as a contemporary "technological hoax." Like The Turk, AMT masked the human labor supporting many digital systems creating the impression of fully automated processes [13]. Another example of this hidden human contribution is ImageNet, a foundational dataset for deep learning, which was compiled through the efforts of over 48,000 contributors from 167 countries via AMT [14]. The scale and profound impact of ImageNet hightlight how, despite advances in automation, much of today's AI progress continues to depend on extensive and often invisible human work behind the scenes.

1.3 Surge of Conversational Agents

The term *artificial intelligence* was formally proposed at a conference at Dartmouth University in 1956 [15]. Since its formal inception, AI development has experienced multiple cycles of intense excitement, often called "hype cycles," followed by periods of stagnation known as "AI winters". The current, third hype cycle has been uniquely accelerated by the release of ChatGPT in November 2022. This conversational, text-based AI quickly became the fastest-growing consumer application in history, attracting one million users within just five days and reaching 200 million weekly active users by 2024 [16]. To put this rapid adoption into perspective, by August 2024, 39% of U.S. adults aged 18–64 had used generative AI, with 32% engaging with it in the previous week [17], a scale of adoption that took years for earlier technologies like computers and the internet to achieve.

Currently, most interactional AI interfaces remain predominantly text-based, however, CAs are widely anticipated to become the next significant breakthrough in AI technology [18]–[20]. These advances promise to fundamentally transform human-machine interaction and open new possibilities across a variety of application domains. Notably, shortly before the submission of this thesis, two major developments in the field were announced.

First, Sam Altman, CEO of OpenAI, together with Jony Ive, the renowned designer of the original iPhone, revealed a collaborative effort focused on creating a new generation of AI-enabled devices designed to empower users to "create all sorts of wonderful things" [21].

Second, Google shared an update on its research prototype, Project Astra, highlighting its potential as an AI assistant capable of functioning as an "AI tutor" to help teenagers solve complex problems, such as math exercises [22].

If this trajectory continues, future CA systems are likely to feature more natural, voice-based interactions, characterized by human-like speech and the ability to sustain meaningful conversations. These systems would also perform tasks seamlessly within the devices they inhabit. This shift is already evident in emerging industry trends, with voice-based CAs increasingly deployed as replacements for help desk personnel and integrated into consumer-facing products such as smartphones.

1.3.1 Children as Emerging Users of CAs

Children are among the fastest adopters of generative AI technologies. For example, a report by the European Parliamentary Research Service found that by 2024, 77% of UK children aged 13–18 had used generative AI tools—more than twice the adoption rate of adults [23]. As these technologies continue to evolve rapidly, children are increasingly growing up in environments shaped by conversational AI (CA) systems.

Voice interaction in particular is a rapidly expanding frontier for children's engagement with AI [18], [20]. This surge in voice-based technologies coincides with a concerning global decline in reading proficiency. According to the OECD's 2022 report, reading scores have dropped by an unprecedented 10 points in several countries since 2018, which is a decline that cannot be fully attributed to the duration of COVID-19 school closures [24]. This suggests that children's modes of accessing and processing information may be shifting toward voice and conversational modalities. Despite this shift, there remains a significant gap in our understanding of how CAs influence children's learning and development.

The growing presence of CAs in children's daily lives is still underexplored, raising important questions about potential mismatches between children's expectations and the actual capabilities of these technologies. More critically, interactions with such systems may shape how children communicate and relate socially, sometimes producing unintended or problematic outcomes. Recognizing these concerns, governments and NGOs worldwide have stressed the urgent need for privacy, safety, and ethical frameworks to guide the responsible design and deployment of AI systems for children [25]–[31].

In everyday life, children already frequently interact with voice interfaces embedded in technologies such as smartphones, smart speakers, and educational applications [32], [33]. As children increasingly engage independently with AI-enabled systems, it becomes essential to understand how these interactions impact their cognitive and social development [34]. However, as we will discuss further in this chapter, research exploring children's preferences and expectations around these interactions remains limited.

Play is a fundamental and universal aspect of childhood across cultures, and provides a key context for how children engage with new technologies [35]. Although research on children's interactions with AI is still emerging, existing studies indicate that children primarily use AI tools for homework assistance and gaming [36]. Moreover, children tend to prefer AI systems that are visually engaging and voice-enabled, rather than relying solely on text-based interfaces [37]. They actively incorporate AI into their creative activities and adapt their learning strategies based on the specific task or domain [38], underscoring the dynamic and context-dependent nature of these interactions.

1.3.2 Perception of AI-generated Voices

The specific voice characteristics for both the adult and child interfaces were selected based on a literature review examining how different AI-generated voices are perceived within HRI contexts. This review helped identify voice features that are likely to influence positive emotions and creativity during play.

Aylett et al. argue that "any system using voice to communicate becomes personified by that voice" [39]. Supporting this, Xu and Warschauer found that children tend to attribute psychological characteristics to CAs even when they lack physical embodiment or mobility [40]. In contrast earlier research on traditional computers, such as studies by Scaife and van Duuren, demonstrate that these devices are often perceived merely as intelligent tools, without associations to psychological or behavioral traits [41], [42].

When considering children's preferences specifically, Hubbard et al. identified three commonly favored voice types in children's suggestions: family voices, robot voices, and character voices [43]. Other studies, including one by Druga et al. indicate that children prefer natural-sounding voices perceived as friendly and sincere, while voices perceived as overly authoritative or "know-it-alls" tend to be disliked [44]. Yuan et al. found that children appreciate personified interfaces, defined as named agents speaking in the first person, though only younger children responded positively to personalization features where the system refers to personal information about the user, e.g. calling user by their name [45].

Another study highlighted the limited availability of child voice options in current systems [43]. Despite the increasing integration of AI in children's lives, few AI frameworks and design guidelines explicitly address voice and conversational elements as critical components of child-AI interaction [46]–[48]. This underscores an urgent need for further child-centered research focused on voice interface design.

Regarding adult preferences, research indicates that higher-quality synthetic voices are generally favored over lower-quality ones [49], with neutral tones being particularly well-received [50]. Preferences also differ by age group: older adults tend to prefer female voices, while younger adults often favor male voices [50], [51].

Given the growing prevalence of human-sounding conversational agents embedded in everyday devices such as smartphones, tablets, and smart speakers [40], we selected an AI-generated, lower-pitched female voice designed to convey a sense of androgyny. This choice aimed to minimize bias arising from gender-based user preferences. We also opted for voices with a calm and slower tempo to facilitate easier comprehension of voice-based instructions during gameplay. Our selection was further guided by our own subjective experience as being a human, choosing a voice that sounded friendly and genuine. Basing on prior research emphasising children's preference for personified agents, we named the voice system we are using in the study "Bobbi," who introduces itself to participants at the beginning of the game.

1.4 Children Learn Creativity Through Play

Play is a universal activity through which children develop a broad range of essential skills, including creativity. Creativity plays a particularly crucial role by fostering flexible thinking and problem-solving abilities, which are vital for navigating complex and ever-changing environments [52].

According to the collaborative report Learning Through Play by LEGO and UNICEF, children engage in and develop different skills through play at various developmental stages. Supporting children's agency and playful learning can take many forms ranging from child-led free play and guided play with adult scaffolding, to rule-based structured games as seen in figure 1.4. Each mode offers unique opportunities for children to take initiative, explore creatively, and cultivate essential skills [1].

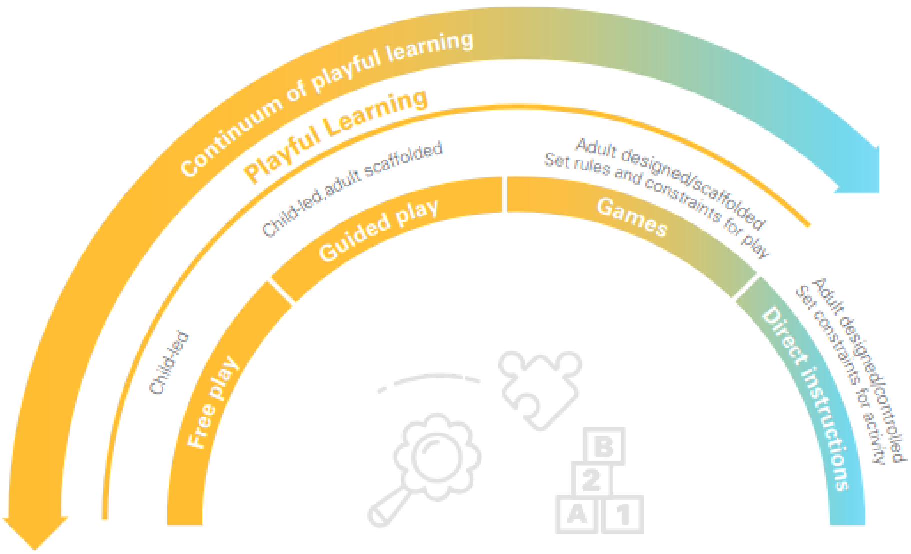

Figure 1.4: Continuum of Playful Learning. Source: UNICEF 2018, adapted from Zosh et al., LEGO Foundation (2017) [53].

Prior research indicates that social robots can effectively enhance creativity. For example, Kahn et al. conducted a study where young adults performed a creative task involving a Zen rock garden either alongside a socially interactive robot (Robovie) or using a PowerPoint interface. Participants interacting with the robot produced significantly more creative outputs and invested more time in the task, demonstrating the robot's capacity to sustain and stimulate creativity [54]. Building on this, Ali et al. explored how social robots support different creativity types: verbal, figural and constructional in children aged 5-10 years old. Their randomised controlled trials revealed that robot interfaces demonstrating creativity themselves particularly enhanced children's verbal creativity, resulting in more original and varied responses [55].

Our study focuses on children aged 9 to 12 years, corresponding to Danish primary school levels 3 to 6, recruited through The LEGO Group's partnerships with international schools in Vejle and Billund, Denmark. These schools teach primarily in English, and children are either native speakers or have learned English as a second language since age 5 or 6. This age group is developmentally appropriate for verbal creativity tasks due to their advancing cognitive and linguistic skills, which enable them to engage in interactive and abstract verbal activities even in a second language [56].

Research shows that creativity expression typically emerges around age 5–6, but often declines during middle childhood (around ages 9–10) due to increasing conformity pressures and standardized learning environments [57]–[59]. Furthermore, children aged 9–12 gain increasing autonomy in interactions with digital technologies and AI-enabled systems, often independently engaging with voice assistants, recommendation systems, and gaming bots [60]. The peer influence growing stronger at this age also shapes how children respond to social cues from embodied AI agents such as robots. Therefore, understanding creativity in this developmental window is both timely and critical.

For this master's thesis, we focus exclusively on assessing verbal creativity. Cross-analysis incorporating emotion data and verbal creativity scoring is planned for future project phases extending through June 2025.

We adopt the definition and criteria for creativity from the Torrance Tests of Creative Thinking (TTCT) [61], defining creativity as the capacity for divergent, original and flexible thinking in problem-solving, idea generation and responding to challenges. Although the TTCT test itself is trademarked and not administered here, its well-established criteria serve as a guiding framework [62] due to several advantages:

- It is widely recognized as a reliable and valid creativity assessment across diverse age groups and educational contexts [62].
- It offers a domain-general view of creativity applicable across tasks and disciplines [63].
- It emphasizes divergent thinking, encouraging multiple original solutions to open-ended problems [62].

- It minimizes cultural biases and subjective evaluator judgment challenges common in creativity research [58].
- It assesses the ability to generate many ideas from a simple prompt within limited timeframes [64].

Building on this foundation, our study investigates how different AI-generated voices (peer-like versus adult-like) impact children's creative expression in a playful setting. The choice to focus on children aged 9 to 12 reflects the importance of this developmental stage, characterized by growing autonomy, linguistic proficiency and a documented decline in creativity due to conformity pressures.

By grounding methodological choices within established theoretical frameworks and developmental considerations, this study aims to deepen understanding of how AI voice interfaces can be ethically and effectively designed to nurture creativity in children. Ultimately, these insights seek to inform the responsible development of CAs that support child-centered learning and creative growth.

For this Master's thesis our focus is solely on the assesment of verbal creativity. The cross analysis of the emotion data and verbal creativity scoring is going to happen on the next phases of the project that will continue til the end of June 2025.

In this study we have adopted the definition and criteria for measuring creativity of used in the Torrance test of Creative Thinking, often referred to as TTCT [61], and define creativity as *the capacity for divergent, original, and flexible thinking in problem-solving, idea generation, and responding to challenges*.

While we are not administering the TTCT test itself due to its trademarked status, we are utilising its well-established and widely recognised criteria [62] to measure creativity. This framework was chosen for several reasons:

1. It has been recognized as one of the most reliable and valid assessments of creativity across various age groups, from elementary school to university students, particularly in educational and developmental contexts [62].
2. It provides a domain-general conception of creativity, meaning it assesses creativity in a broad, generalized way that can be applied across a variety of domains and tasks, rather than being specific to one domain or expertise. [63].
3. It is particularly valued for its ability to assess divergent thinking, the generation of multiple original solutions to open-ended problems [62].
4. Its design minimizes some of the challenges in creativity research, such as biases caused by cultural influences, and subjective judgements of the evaluators [58].

5. Additionally, it focuses on the capacity to generate many ideas from a simple starting point within a limited timeframe [64].

Building on this understanding, our study specifically investigates how different AI-generated voicess, specifically peer-like versus adult-like voices, can impact children's creative expression within a playful context. We focus on children aged 9-12 years old, a critical developmental period marked by increasing autonomy, linguistic proficiency, and notable shifts in creativity. This age group is especially relevant because research suggests creativity often declines during middle childhood due to rising conformity pressures.

In response to these challenges, this study approaches creativity with sensitivity to context and carefully grounding methodological choices within relevant theoretical frameworks. By doing so, we hope to advance the understanding of how AI interfaces can be designed to nurture creativity in children, ultimately informing more ethical and responsible CA development.

1.5 Equality Supports Children's Agency

According to the LEGO Foundation and UNICEF framework, agency is defined as the ability to make meaningful choices and influence one's environment [1]. Moreover, numerous studies have highlighted the important role that peer relationships play in fostering creativity in children [65]–[67].

L. Vygotsky, a foundational theorist in developmental and social psychology, emphasised that children engage differently with adults and peers, each type of interaction supporting distinct cognitive processes [68]. Adult-child interactions tend to be more structured and goal-oriented, where adults provide guidance, direct instruction, and support for tasks that lie just beyond the child's current competence, a concept known as *the Zone of Proximal Development*. In contrast, peer interactions are often less structured and more exploratory, allowing children to collaborate, negotiate and co-construct knowledge.

Supporting this view, Gurdal and Sorbring [69] examined children's perceptions of their relationships with adults and peers. Their findings revealed that 10 year old children reported the highest levels of personal agency in peer relationships, which is important factor linked to learning and creativity, followed by relationships with parents, with the lowest sense of agency reported in interactions with teachers.

Building on this evidence, our hypothesis is that peer-like contexts, in which children feel more autonomous and less constrained, better support creative expression where children are more likely to freely explore imaginative ideas. Conversely adult-led interactions tend to promote convergent thinking, focusing on finding correct answers and following explicit instructions.

Therefore, based on research into children's agency across relational contexts, we hypothesize that children will express more imaginative and unconventional ideas when interacting with a peer-like voice interface, compared to an adult-like voice interface. The peer-like voice is expected to encourage more playful and elaborated idea generation, while the adult-like voice may elicit more systematic, less elaborated responses. This study investigates this hypothesis by comparing children's verbal creativity during a creative task delivered via two voice interface conditions: one with a child-like voice and the other with an adult-like voice. Our goal is to assess whether differences emerge in the quantity, originality and elaboration of children's responses between these conditions.

> "Today's children will redefine the scope and shape of the playing field for social relations in the future. Because they are the first generation to grow up with this new paradigm, it is essential that we observe and document their experiences."
> - The Third Culture, Sherry Turkle [70]

1.6 Structure Overview

The structure of the thesis is as follows:

Chapter 2: Problem Analysis critically examines how creativity is commonly defined and measured in interdisciplinary research. We highlight the disconnect between creativity as a socially constructed phenomenon and its frequent operationalisation as a fixed, decontextualised trait. Drawing on the SSK we show how standardised creativity metrics may introduce bias and oversimplification. Lastly, we present the problem statement and research questions focused on.

Chapter 3:Methodology details our interdisciplinary research design.

Chapter 4: Results presents data-analysis and preliminary results from the first phase of the study with adult participants.

Chapter 5: Discussion reflects on the preliminary findings and their broader implications. We critique the biases in conventional methodologies and emphasise the importance of integrating children's perspectives in AI design. Methodological and ethical limitations are explored and we conclude with a forward-looking discussion on future research directions.

Chapter 2

Problem Analysis

Creativity is inherently difficult to study. It does not follow fixed rules or predictable pathways, and its value is deeply context-dependent. Despite this, research across disciplines often reduces creativity to standardised, measurable traits, creating a gap between how creativity is understood and how it is studied. This chapter explores this tension through the lens of HRI and broader technological research, asking: what are the risks of treating creativity as a universal and quantifiable concept?

Moreover, creativity often emerges from the synthesis of ideas across different domains — a process that is hard to trace, quantify, or replicate. Individuals may combine concepts in unexpected ways, draw on intuition or unconscious thought, or rely on contextual knowledge that is invisible to the observer. As a result, studying creativity often requires interdisciplinary approaches and flexible methodologies that can account for its complexity and variability.

In short, creativity resists reduction into simple variables. Its richness lies in its unpredictability — which is exactly what makes it both fascinating and challenging to research.

2.1 Creativity in the Context of HRI

In many technological and interdisciplinary fields such as HRI, creativity is increasingly integrated into study designs (e.g. [71] [72] [54] [73]).

While there is growing academic interest in the intersection of creativity, AI, and children, it also raises a critical concern: as creativity becomes more integrated into HRI research, it risks being treated as a fixed and universal concept. Without critical engagement with the theoretical complexities of creativity, these studies may overlook how their definitions and measurements reflect culturally specific assumptions, especially when applied to diverse and vulnerable populations such as children. This echoes similar challenges seen in related concepts like engagement, where diverse definitions across disciplines complicate cohesive understanding and application [74]. Without clearer conceptual grounding, creativity risks being treated as a universal and easily transferable construct, which can weaken the theoretical coherence of findings across studies.

A widely recognized challenge in creativity research is the difficulty of reaching consensus on a single definition and the related problem of how to effectively measure it [57], [62], [75]–[77]. This issue is commonly referred to as the criterion problem. It describes the methodological difficulty in quantifying and assessing complex, abstract human experiences such as creativity specifically determining which criteria genuinely represent creativity and how these criteria can be measured in a meaningful and reliable way [57]. Because creativity is inherently subjective and elusive, existing metrics may fail to capture its full essence, potentially leading to misleading or incomplete conclusions that do not accurately reflect the phenomenon under study.

2.2 Historical and Disciplinary Roots of Creativity Research

Although creativity is a familiar term that people use and understand with ease in everyday conversations, if asked, people would still define it in various different ways. For some it can mean "getting unstuck when you are stuck”, or "regression to a more child-like and playful state of mind", or "being a wonderful artist, musician, writer, or inventor" - this offers only a small glimpse into the numerous ways people describe what creativity is, as compiled by D. J. Treffinger. [78] Despite different individual viewpoints, in Western culture - typically understood today as encompassing Europe and North America [79] - creativity is generally viewed as a positive attribute. [80]

Contemporary understanding of creativity has been historically shaped by theories, models, and experiments developed by scholars working in Western academic institutions. The conceptual roots of creativity lie in cognitive psychology [81], particularly in its early work attempting to understand the mental processes behind creativity [82]. Cognitive psychology emerged predominantly in Western contexts during the mid twentieth century as the emerging fields such as linguistics, neuroscience, and computer science began to shift scientific attention toward the study of mind. This broader shift in paradigms prompted also psychology to move towards studying internal mental processes. This shift from behavioristic psychology marked the beginning of what would later be known as the cognitive revolution. [83]

The Western emphasis on individualism, originality and autonomy [80] has influenced how creativity has been theoretisised and studied, shaping both the focus of research and the characteristics most often associated with being creative [84]. As a result, the early research on creativity was often focused on how creativity happens [82], by focusing primarily on the cognitive processes and personality traits associated with creative individuals [62], [85], [86].

In one of the textbooks that many psychology student read during their studies, *Cognitive Psychology - Student's Handbook* (2005), by the authors M. W. Eysenck and M. T. Keane, the chapter about creativity stays true to its historical foundations and introduces mostly different internal models of creative processes such as Walla's (1926) classification model and dual-process theories. Eysenck, as well as other early cognitive psychologists, have noted that many cognitive models of creativity often fall short on identifying the precise internal mechanisms behind creativity, and rather tend to be more descriptive in nature. Despite of this, the field continues to offer valuable insights into individual differences and mental strategies that support creative thought, and have provided useful concepts to understanding creativity such as divergent and convergent thinking, mental shortcuts, and analogical resoning. [82]

Over time, creativity researchers increasingly argued that emphasizing the outcomes of the creative process provides a more practical and research-friendlier framework than focusing solely on the internal cognitive processes [57], [75], [87]. This shift in perspective happened timewise as the academic field was going through yet another remarcable shift in paradigm known as *the crisis debate of social psychology*. This refers to a period, particularly in the 1970s, when social psychology faced internal criticism about its methods, relevance, and lack of attention to social context and real-world issues. Scholars like R. Harré, H. Tajfel, and others questioned the heavy reliance on laboratory experiments and the dominance of positivist, individual-focused approaches, and have critisized for lacking cultural sensitivity, and overgeneralising of findings that are actually shaped by Western cultural and educational contexts. The crisis led to a growing emphasis on developing approaches within the discipline that were more qualitative, critical, and sensitive to cultural context. [88]

Alongside with the cognitive definitions and models started to develope a branch of creativity research that increasingly complements internal perspectives with sociocultural, and interactional views [57], [75]–[77], [89] where creativity is often broadly defined as the ability to generate responses that are both original and useful [82], [90], [91]. These newer frameworks abandon the idea of one universal definition, and instead they recognise that what is considered creative in one domain, may differ highly in other areas [58], [62], [92], and have increasingly revealed how cultural perspectives shape conceptions of creativity, particularly highlighting the contrasting ways creativity is understood and appreciated in Eastern versus Western cultural traditions [93], [94], [95], [96], [84].

Conceptualizations of creativity vary across cultures, reflecting differing values and emphases. Western cultures tend to emphasize creativity as originality and the production of final, novel outcomes, whereas Eastern cultures focus more on skill development and the creative process itself [84]. Cultural differences in the valuation of originality and usefulness can lead to divergent interpretations of what constitutes creative behavior. In many East Asian cultures, the usefulness or practicality of an idea is prioritized over its originality [97]. Furthermore, uniqueness, which is one central component of divergent thinkgn that is considered related to creativity, is often understood as deviation which may be less favored in East Asian contexts due to cultural norms emphasizing harmony and conformity [94]. This cultural discrepancy highlights the need to consider cultural frameworks when evaluating creativity, as measures developed primarily in Western contexts may not fully capture the nuances of creativity as experienced and valued in other cultural settings.

Despite the wide range of definitions, most studies on creativity nowadays focus on two core aspects *fluency* and *originality* which are commonly used to assess creativity and can be applied to different contexts and settings. Fluency refers to the ability of an individual to produce a large number of ideas quickly within the given timeframe and context. In contrast, originality focuses on how unique or novel these ideas are when compared to those generated by others within the same context. Together these factors help measure creativity as they reflect both *the quantity* and *the quality* of ideas. However, even here it depends on the research goals and study design how well fitted these measurements are, and how they are best applied and measured. [98], [99]

However, despite the long history of creativity research, providing a systematic and universally accepted academic definition still is an ongoing challenge. This challenge is well documented in the scientific literature reviews (e.g. [62], [78]). A good example that highlights the complexity of defining creativity is D. J. Treffinger's (1996) effort to capture this diversity. Treffinger compiled a comprehensive compendium of creativity definitions from 112 scientific publications publishen between the years 1926-1994, and on their review of these publications documented over 100 different definitions for creativity [78].

For the purpose of this thesis we leveraged natural language processing by processing the abstracts of all 112 publications included in Treffinger's compendium through ChatGPT. We asked the AI with to extract overlapping themes and recurring definitional elements to identify areas of conceptual common grounds, and it retrieved us the following list including dimensions such as *novelty, originality, usefulness* and *transformation*, also emphasing the *contextual nature* of creativity:

> "Nearly all definitions emphasise novelty—the idea that creativity involves producing something new, original, or unexpected. At the same time, this novelty is almost always paired with value or appropriateness. Another shared theme is the idea of transformation—not only producing something entirely new, but reconfiguring existing elements to generate new insights or functions. This points to creativity as a dy-

namic process rather than a static product. More recent perspectives also highlight the social and cultural embeddedness of creativity. It is not just an individual trait or ability, but something that happens in relation to others, to tools, to norms, and to the broader cultural environment." (ChatGPT, 2025)

This AI-generated synthesis aligns with our own literature review of the history of creativity research.

2.3 Challenges in Measuring Creativity

As a direct consequence of creativity lacking a single universal definition, there is also no standardized or widely agreed-upon method for measuring it [62]. This absence of consensus makes designing measurement methods that are both accurate and adaptable across diverse contexts a central and ongoing challenge in creativity research [100].

Creativity is currently assessed using a wide range of methods, each grounded in distinct conceptual frameworks and designed to capture specific dimensions of creativity - that ideally align with the particular context of the study [62], [85]–[87]. These methods vary in their effectiveness: some provide strong validity for narrowly defined purposes, while others offer greater flexibility and applicability across broader contexts. Crucially, no single method comprehensively measures all facets of creativity, given its inherently multifaceted nature.

Figure 2.3 illustrates several creativity assessment tools, highlighting their unique strengths and limitations. Due to this inherent diversity, methodological triangulation, combining multiple measurement approaches, is often recommended to develop a richer, more nuanced understanding of creativity [62], [101].

This variability in definitions and measurement approaches underscores the importance of carefully considering the contextual factors and theoretical assumptions behind each method. Without such reflection, comparing results across studies or drawing generalized conclusions becomes problematic, as each approach embodies different values, priorities, and interpretations of creativity [62]. Furthermore, creativity assessments frequently depend on subjective judgment, which complicates efforts to isolate and quantify creativity in purely objective terms. Consequently, the pursuit of a single, standardized test or universal "one fits all" approach that can be reliably applied across diverse contexts remains an elusive goal [58].

While a comprehensive review of all creativity measures in related research is beyond the scope of this thesis, this section will present illustrative examples to highlight some approaches to demonstarate this variety.

A notable perspective among creativity researchers suggests that creativity is

Approach	Focus	Conception of Creativity	Instruments	Advantages	Weaknesses
Process	Creative processes or skills associated with creativity	Domain-general	WKCT (Wallach & Kogan, 1965); TTCT (Torrance, 1966, 2008); SOI (Guilford, 1967); CAP (Williams, 1980)	Widespread utility High reliability Standardized criteria for interpreting scores	Limited scope of measurement conflicting evidence for validity Bias due to scoring and sample size
Person	Personality traits or creative achievements	Domain-general or domain-specific	HDYT (Davis & Subkoviak, 1975); CPS (Gough, 1979); HCAY (Raudsepp, 1981); CBI (Hocevar, 1979c); CAQ (Carson et al., 2005); BICB (Batey, 2007)	Ease of use High reliability Standardized criteria for interpreting scores	Limited scope of measurement Low validity of self-reports Bias due to self-reporting Neglect of differences in creative personality across domains Low sensitivity to training Skewed scores
Product	Creative products	Domain-specific	CAT (Amabile, 1982)	Similar to evaluating creativity in real-life High reliability High validity	Limited scope of measurement Difficulty in selecting judges Bias due to judges Expensive and time-consuming Lack of standardized criteria
Press	Work environment or climate	Domain-general	SSSI (Siegel & Kaemmerer, 1978); CUCEI (Fraser et al., 1986); WES (Moos, 1986); WEI (Amabile & Gryskiewicz, 1989); KEYS (Amabile et al., 1996); TCI (Anderson & West, 1998); SOQ (Isaksen et al., 1999)	Explore whether a work environment is supportive or inhibitive of creativity Evaluate the environmental improvement attempts and corrective actions	Limited scope of measurement Lack of research-based evidence Debate about “climate” meaning and measurement level Individual differences in conception of climate

Figure 2.1: Measuring creativity [62]

best assessed through outcome-based methods, that is evaluating the creative output. Whereas some researchers argue that this is best done through two complementary measures: the quantity of creative accomplishments and the significance of the individual’s most outstanding achievement [87], others emphasize tests of divergent thinking, which assess an individual’s capacity to generate multiple novel ideas in response to a stimulus. These tests are widely regarded as reliable and objective indicators of creative potential [91], [102], providing a useful complement to assessments focused on actual creative outputs.

However, Rákóczi and Szitó have critically addressed the issue of test bias in the identification of gifted students, particularly when standardized creativity assessments are used with socially disadvantaged populations [103]. Their study was based on the MONDALK Test, which is a divergent thinking assessment developed in Hungary and demonstrates that conventional cut-off points (e.g. the 90th percentile) may systematically under identify high-ability students from marginalised backgrounds. Instead, they argue for adjusting the threshold to the 80th percentile to better capture the originality and quality of ideas among disadvantaged students (ibid.). This finding highlight the importance of critically evaluating the epistemological assumptions embedded in measurement tools and reveals how even well constructed instruments can reinforce structural inequalities if their application does not critically consider who the tests were designed for and who they may exclude.

In addition, Runco et al. demonstrated that in two seperate studies that divergent thinking tests, that are commonly used to estimate creativity, are significantly influenced by an individual's prior personal and social experiences. Their studies showed that originality and fluency scores were not purely measures of creative ability but were up to 65% explained by prior experiences. This experiential bias suggests that DT tests may unintentionally favor individuals with broader or more enriched life backgrounds, and therefore limiting their fairness in evaluating creativity across diverse populations. [99], [104].

E. P. Torrance, the creator of the Torrance Test for Creative Thinking, commonly referred to as TTCT, distinguished their test from other creativity and divergent thinking assessments not only through their specific definition of creativity but also by designing the TTCT to be engaging, easy to administer, and suitable for diverse populations and cultural backgrounds [105]. Creativity according to TTCT is defined as the capacity for divergent, original, and flexible thinking in problem-solving, idea generation, and responding to challenges. TTCT further provides quantifiable criteria to observing these creative capacities. It emphasises following dimensions of creativity [61]:

- **Fluency:** The ability to produce a large number of ideas or solutions.
- **Originality:** The uniqueness and novelty of generated ideas.
- **Flexibility:** The ability to shift perspectives and adapt to new ways of thinking.
- **Elaboration:** The capacity to refine and expand on ideas by adding detail and depth.
- **Resistance to premature closure:** The ability to withhold judgment and allow time for original ideas to develop.

Although the TTCT is widely considered a reliable tool for measuring creativity across different domains, contexts and cultures [62], it is not without limitations. Like other standardized assessments such as the Wallach and Kogan tests, the TTCT was developed within specific cultural and educational settings. These backgrounds have shaped the underlying ideas about what creativity is and how it should be measured [61], [64]. While these tools aim to be broadly applicable, their definitions of creativity may not fully reflect the diversity of how creativity is understood and expressed in different cultural contexts.

The TTCT focuses on tasks that ask individuals to generate many original ideas, placing importance on the production of visible outcomes. This reflects a Western perspective on creativity, where being creative is often associated with producing something new and observable [106]. These values influence the structure and focus of the test itself, highlighting certain types of creative expression over others. Therefore, it is not always whether such approaches align with how creativity is defined and valued in other parts of the world. In some cultures, creativity is connected to collective practices and problem-solving within social contexts, rather than individual output or originality [106] [94].

To address the persistent criterion problem in creativity research, the challenge of defining and reliably measuring creativity T. M. Amabile, a prominent American psychologist and researcher known for their influential work on creativity, emphasized the importance of confronting this issue directly and systematically [57]. In addition, Blamires et al. proposed a structured set of reflective questions that researchers should carefully consider within the specific context of their studies on creativity. These guiding questions are intended to prompt critical thinking about what constitutes creativity in a given domain, how it is recognized, and how it can be meaningfully assessed [107]. These questions are as follows:

- How should creativity be defined in the context of this study?
- Which specific aspect of it need to be valued and measured?
- And, which tools and methods are most appropriate for assessing these aspects? (ibid.)

If these foundational questions are not addressed, researchers risk providing weak or superficial justifications for the choice of metrics. Without a clear and contextually grounded definition of creativity, it becomes easy to rely on inadequate or mismatched assessment tools believing they capture creative potential when, in fact, they may be measuring something entirely different. This can lead to misleading conclusions.

For instance, in using the TTCT as a reference point for designing creative tasks in HRI, we have been attentive to its underlying conceptual foundations. Considerable time was dedicated to ensuring that the study design is theoretically grounded, aligning both the tasks and evaluation criteria with established understandings of creativity. As the developer of the TTCT test, E. P. Torrance themselves said:

> "For any kind of use of the TTCT, it is important that the users have at least basic knowledge of the rationale for the tasks or activities, and be familiar with the concepts of creative thinking that underlie the instrument" [108].

2.4 Problem Statement

This thesis addresses the problem that current creativity research often employs standardised measures and simplified definitions that inadequatly capture creativity's complex, context-dependent nature.

- In research that has adopted creativity as a concept away from its original roots, there is a lack of critical reflection on the epistemological assumptions underlying creativity measures, and how social and cultural contexts influence what is considered creative.
- The widespread use of decontextualised tools risks misrepresenting creativity and may exclude or marginalise diverse forms of creative expression.
- This issue is particularly acute in research involving vulnerable populations (e.g. children) or emerging technologies, where creativity may manifest in novel or unconventional ways.
- There is a need for more contextually sensitive, interdisciplinary approaches that recognise creativity as a socially constructed and dynamic phenomenon.
- In technical research creativity is frequently operationalised through quantitative metrics, while the sociocultural assumptions that are embedded in these measurements are rarely examined.

This thesis aims to fill this gap by critically analysing existing creativity measurement practices through the lens of SSK and interdisciplinary research, ultimately proposing more nuanced frameworks for understanding and assessing creativity.

2.5 Research Questions

How can context-sensitive, interdisciplinary approaches improve the way creativity is evaluated in child-AI interactions, particularly when mediated by conversational agents (CAs) with different perceived age characteristics?

How do adult participants respond creatively to child-like versus adult-like CAs in a structured verbal task?

Chapter 3

Methodology

This chapter presents the methodological approach for investigating the effects of perceived age in voice interface variations on creativity in playful interaction. A mixed-methods two-phase study was conducted, involving adult participants interacting with a turn-taking Droodles game using synthesized child and adult voices. The interface was developed as a local web application with voice synthesis powered by the Narakeet API. Data collection focused on verbal creativity measures based on the TTCT model, with analysis combining quantitative and qualitative techniques. The study design was informed by interdisciplinary theoretical frameworks including Sociology of Scientific Knowledge, developmental psychology, and learning through play paradigms. The following sections detail the study design, materials, procedures, data collection, analysis, and conceptual grounding.

3.1 Study Design and Participants

A two-phase study was conducted at the AAU Robotics Lab.

- **Pilot Phase:** February 2025, five adult participants tested an early version of the voice interface to identify technical and interaction issues.
- **Main Phase:** May 2025, 24 adult participants recruited from university communities and stakeholder groups participated in the finalised interface testing at AAU Robotics Lab. Participants were primarily recruited from university communities and relevant stakeholder groups, with all participants providing informed consent in accordance with ethical guidelines.
- **Next Phase:** June 2025, Future phases will include children tested at LEGO in June 2025 to explore age-related differences in creativity in response to voice interfaces.

3.1.1 Participant profile

- 9-12 year old English speaking children

3.1.2 Study conditions

- Child Voice Interface: system used a childlike synthesized voice
- Adult Voice Interface: system used an adult-like synthesized voice

Quasi-random group assignment

Rather than differentiating participants by native vs. non-native English proficiency, participants are quasi-randomly distributed to two different study groups based on their age and gender. We focused on age-related cognitive and linguistic development as the primary factor influencing their ability to participate meaningfully in the verbal creativity task. This approach acknowledges that a 9-year-old and a 12-year-old child may differ substantially in verbal fluency, self-expression, and social reasoning skills [56], [68].

Additionally, research in Denmark highlights gender-based differences in English language development among 9–12-year-old children. Boys often outperform girls in certain language competencies, a phenomenon partly attributed to greater exposure to English through online video games, which provide immersive, interactive, and often peer-based language experiences [56]. Based on this, we considered gender as a potentially significant variable in our study design, as it may influence both language performance and interactional patterns with AI in play-based learning environments.

3.1.3 Game design

The game interface we designed for the study is based on Droodles, that are simple, abstract doodles or sketchs that is deliberately ambiguous and open to interpretation. The term is a blend of words doodle and riddle. Droodles are typically used to spark creativity and humor, as participants are encouraged to come up with imaginative explanations or stories about what the drawing represents. They were popularized by Roger Price in the 1950s, especially through their book "Droodles" [109], which presented these minimal drawings with humorous or surprising captions. Droodles have since been used in previous creativity studies as prompts for creative expression due to their abstract nature that invites to open ended idea production [110], [111], [71], [112], [113], [114].

The TTCT, which served as a methodological foundation for defining creativity and measuring it in this study, emphasises creativity as a capacity to generate multiple and diverse ideas from a simple prompt within a constrained time frame

Figure 3.1: Examples of Droodles. Source: Droodle Compedium, Tallfellow Press [115]

[64]. In alignment with this principle, the play activity designed for the study was intended to mirror the TTCT's emphasis on fluency and divergent thinking under time pressure, affording spontaneous and creative expression within certain restrictions.

The game interface was developed as a local web application using HTML, CSS, and JavaScript, served via a Flask server on a local machine. Voice synthesis was powered by the Narakeet API, which generated natural-sounding child and adult voices. Voice clips were pre-generated to ensure consistent delivery.

Two screens were used: one for the participant's view and one for the researcher's control and monitoring, including real-time observation and tracking of participants' emotional responses.

3.1.4 Turn sequence

Each player, both the participant and the voice system, took five turns, each limited to 30 seconds. During each turn, a Droodle image was displayed on a screen in front of the participant, who was then asked to guess what the picture could represent. Each participant took five turns of up to 30 seconds each to verbally interpret the Droodle image shown.

The voice interface responded with pre-scripted creative narratives modeling:

- Fluency (multiple alternative ideas)
- Elaboration (rich descriptive detail)

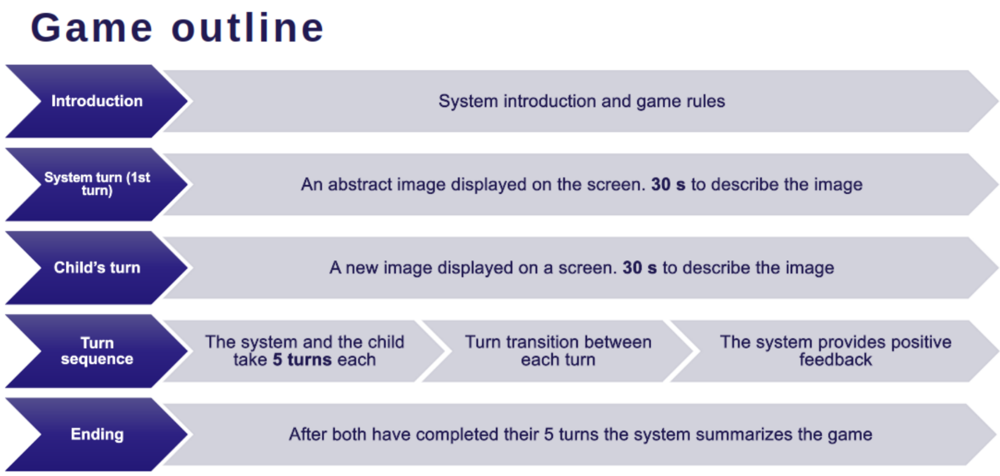

Figure 3.2: The game setup and turn sequences. Source: "Learning Through Play With AI" study design, M. Alaoja 2025

- Originality (unexpected imaginative concepts)

3.1.5 Technical design of the interface

The technical design of the game interface was created by our AAU team member Suzy Choi. The voices were synthesised using the Narakeet API, which uses generative AI to produce lifelike speech in both child and adult voice styles. All voice clips were pre-generated prior to the sessions to ensure consistency. The interface itself was implemented as a local web application using HTML, CSS, and JavaScript, and was served locally via a Flask server.

3.2 Study Setup

The game interface and voice system were operated using a Wizard of Oz (WoZ) setup, in which two members of the research team manually controlled the system from within the same room as the participant but remaining out of sight. This method was intentionally selected to ensure child safety and data privacy by avoiding direct interaction between participants and a generative large language model (LLM), reducing risks with unpredictable or inappropriate AI-generated responses.

During the game, the participant sat facing a screen where the game content was displayed. Meanwhile, the researchers controlled the flow of the interaction and monitored the participant's verbal and non-verbal responses in real time.

Figure 3.3: A pilot testing of the game interface at the AAU in February 2025

These emotional and behavioral reactions were manually logged and will later be compared with emotion-tracking outputs from the AI system.

3.3 Data

3.3.1 Data collection

- Verbal responses transcribed for analysis of fluency, elaboration, and originality
- Emotional expression observed and tracked by researchers
- Emotion tracking from video by AI

In this study, we are recording participants' answers during the game to assess their verbal creativity expressions. The study also employs a multimodal approach to emotion tracking, combining facial expression and audio analysis to measure participants' emotional responses during interaction with the AI voice interfaces. The emotional responses of participants are going to cross-analysed together with the creativity scoring for gaining a more multifaceted understanding of the effects of the interfaces.

- **Vision (Facial Expression):** Multiple emotion recognition models are evaluated using the DeepFace framework, including exploration of large vision-language models like VLM (LLava). Due to most models being trained on adult faces, plans include fine-tuning with children's data if available.

- **Audio Emotion Analysis:** Speech emotion is analyzed with the FunASR framework and emotion2vec model, comparing audio-only with multimodal models integrating speech content for enhanced accuracy.
- **Multimodal Integration:** After evaluating individual modalities, the project aims to combine facial, audio, and text data for comprehensive emotion recognition.
- **Engagement Metrics:** Additional behavioral data such as body posture, reaction times, and conversation frequency are recorded to assess engagement.
- **Data Collection Hardware:** Two cameras (Logitech BRIO 4K for frontal facial capture and Microsoft Azure Kinect DK for posture and scene analysis) and a Blue Mic Yeti USB microphone are used to capture video and audio data.

3.3.2 Data analysis

- Transcription of the audio files
- Dissecting the data according to TTCT model (involves coders qualitative input and evaluation)
- Descriptive statistics
- Statistical methods to compare creativity scores between the child and adult voice interface conditions.

In this study we are adapting the creativity metrics and definition from the Torrance Test of Creative thinking, and following commonly used metrics of fluency, originality and elaboration to measure participant's creative expressions.

3.4 Ethical Considerations

The project is commited following Aalborg University guidelines for child protection policy [116], and main principles and procedures. Ethical approval was obtained from Aalborg University's Research Ethics Committee in April 2025 to ensure full compliance with regulations governing research with minor participants. In addition special attention in study design was paid to ensure that children participating in the test have a option to safely exit the game if they wished, as well as creating informed consent forms for guardians of children, while also being sensitive to what kids themselves want.

3.5 Limitations

The study was conducted in a controlled laboratory environment, while it is useful for maintaining internal validity, may not fully reflect the dynamics of naturalistic play. Real-world contexts such as home or classroom environments might elicit different patterns of interaction, especially in children. Therefore the ecological validity of the findings may be limited.

The sample of adult participants primarily consisted of individuals from university communities and affiliated stakeholder networks. While in general this kind of sampling strategy may introduce bias and limit the generalisability of results across more diverse populations, in our case it is justified bacause of our framing the study in SSK we are taking into account the context dependent nature of creativity we are not aiming to produce generalisations over different contexts, but to see if inside of this context of samples there are differences between participants answers depending on the voice interface used. The results pointing towards possible correlational factor between the responses and whether a child-voice or adult-voice interface was used can be taken into account in other future research and design projects as a factor that can possible be important to consider in that new context also. Similarly, the child study, planned for a next phase of a project, may face similar demographic constraints.

The pilot phase was limited in scale, with only five participants, and primarily served interface refinement purposes rather than data collection for analysis.

From a methodological perspective, the study was shaped by interdisciplinary collaboration. While this enriched the design and analysis, it also introduced challenges in aligning epistemological assumptions and methodological preferences across fields. Balancing technical rigor with socio-cultural reflexivity remains an ongoing negotiation in interdisciplinary research.

Finally, creativity itself is a context-dependent and culturally mediated construct. Despite efforts to design a contextually valid and inclusive study, the evaluation of creativity remains subject to interpretive variability. While structured rubrics were developed based on established creativity assessments, some degree of subjectivity in scoring remains inevitable.

These limitations do not undermine the contributions of the study but rather highlight the complexity of researching human-AI interaction within creative contexts. They also point toward directions for future research, including more diverse sampling, real-world deployment, and the integration of embodied robotic systems.

Chapter 4

Results

4.1 Data Description

One participant's data was excluded from the analysis due to incomplete audio recording, so our dataset included entries from 23 participants, 11 interacting with a child voice interface, and 12 interacting with an adult voice interface. Data included the total of 333 ideas created by participants, out of which 237 were elaborated further.

4.1.1 Fluency and Elaboration

Table 4.1 presents the average number of ideas generated per picture by participants in each study condition. Across all five pictures, participants using the adult-voice interface consistently produced slightly more ideas.

Average Fluency	Picture 1	Picture 2	Picture 3	Picture 4	Picture 5
Child Voice Interface	2.583	2.417	3.083	2.750	2.750
Adult Voice Interface	3.000	3.273	3.818	2.818	3.455

Differences and their consistent direction across all pictures may suggests a stable trend in favor of the adult-voice condition in terms of fluency, the creation of multiple ideas to stimuli in a restricted short time frame.

This pattern is visualised in Figure 4.1, which illustrates the average number of ideas per picture for both groups. The figure highlights how the adult-voice condition resulted in marginally higher number of ideas for each picture, suggesting that the adult voice may have better afforded procuction on multiple ideas.

Figure 4.2 visualises this consistent trend of participants in adult-voice interface settings producing in average more ideas per picture.

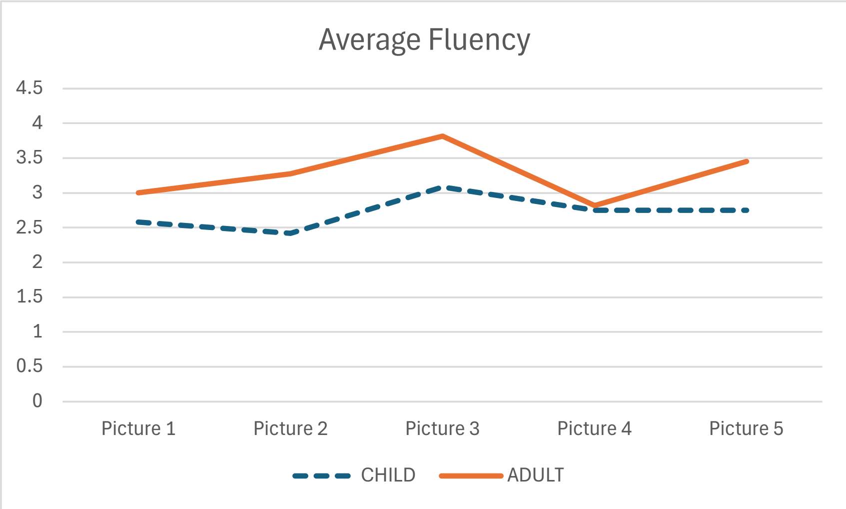

Figure 4.1: Average fluency per picture in two study conditions

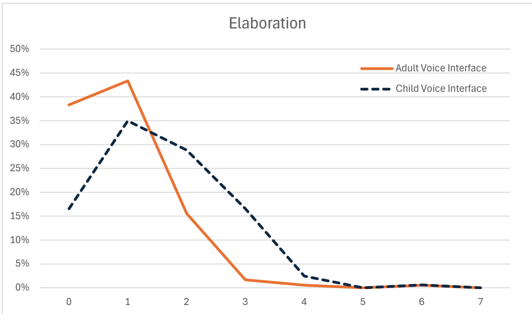

Figure 4.2: Enter Caption

Each instance of elaborating and idea in the data has values between 0 and , referring to the amount of added detail and elaboration to each idea that was presented, where elaboration value 0 rfers to ideas that were not elaborated at all, elaboration value 1 refers to instances where a participant presented an idea which included one dimension of additional detail added to it. Examples of elaboration values from data to demonstrate this are presented in table 4.3.

4.1.2 Distribution of Elaboration

The distribution of elaboration instances across the two voice interface conditions (Table 4.2) provides further insight into how the design of the voice interface influenced participant behavior.

The child-voice condition demonstrated a lower proportion of zero-elaboration instances (18%) compared to the adult-voice condition (38%), indicating that partic-

Table 4.1: Examples of Elaboration Values

Elaboration Value	Description	Examples
0	Idea presented without additional detail.	• *"This can be a speaker"* (Participant 010, Picture 4) • *"Well, it looks a lot like a throttle valve"* (Participant 014, Picture 2)
1	Idea presented with one dimension of added detail.	• *"Bird legs sticking up out of a hole"* (Participant 004, Picture 5) • *"It could be a wrist watch on someone's arm, maybe?"* (Participant 008, Picture 4)
2	Idea presented with two dimensions of added detail.	• *"It could also be two trees. That are coming up from the ground. Taking their look into the world."* (participant 015, picture 5) • *"a person with a very big hat walking up the stairs."* (participant 002, picture 4)
3 (...)	Idea presented with three dimensions of added detail.	• *"some kind of ornament placed on a ladder. Maybe in the garden or something like that. It could probably work as a sun clock of some sort."* (participant 004, picture 4)
6	Idea presented with six dimension of added detail.	• *"A stone, on a field. And the stone has two arms. And it's trying to run away. But has no feet. So it's kind of stuck. Which is kind of sad for the poor stone."* (participant 015, picture 5)

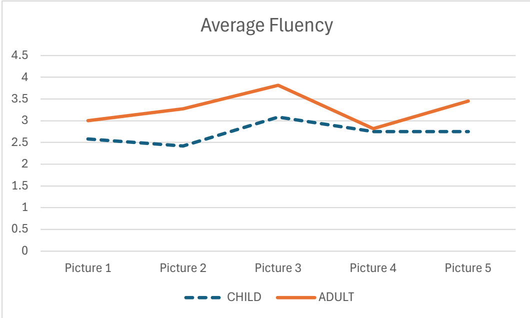

Figure 4.3: new

	Adult Voice Interface		Child Voice Interface	
Elaboration	Occurences	%	Occurences	%
0	69	38 %	27	18 %
1	78	43 %	55	36 %
2	28	16 %	44	29 %
3	3	2 %	23	15 %
4	1	1 %	3	2 %
5	0	0 %	1	1 %
6	1	1 %	1	1 %
Total	180	100 %	153	100 %
Sum of all Elaborations		153		230
Average		0,85		1,50
Standard Deviation		0,88		1,08

Table 4.2: The distribution of elaboration in the two study conditions

ipants were less likely to leave an idea unexplained or minimally described when interacting with the child-voice interface. In addition, the child-voice group exhibited a greater amount of higher elaboration values, with the total of 72 (48%) of all the ideas (n=153) they produces containing two or more additional detail to them, compared to the total amount of 33 (20%) in the adult-voice condition (n=180). This shift toward higher elaboration in the child-voice group aligns with the observed difference in average elaboration (mean 0.85 vs. 1.50) per idea and suggests that the child-voice interface was more effective in encouraging participants to expand their ideas with additional detail.

Furthermore, Figure 4.3 visualises the average number of elaboration instances, with one or more added details to the original idea, per picture in both study conditions.

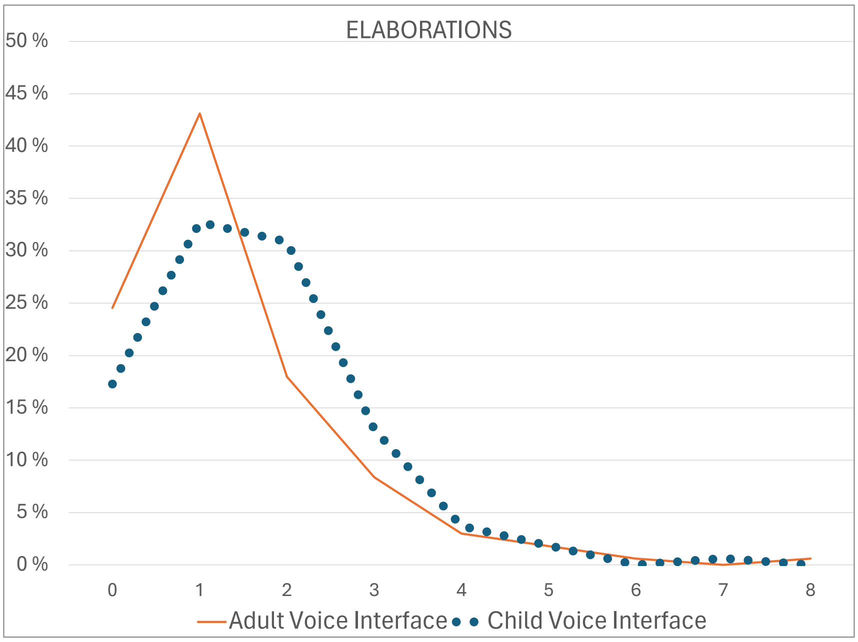

Figure 4.4: Percentual distribution of elaboration instances in the two study conditions

4.2 Summary and Conclusions

The findings in this study points towards the notion that the perceived age of AI voices in CAs can significantly influence the quantity and the quality of participants creative responses. When interacting with an adult-voiced interface, the participants produced a higher number of ideas, suggesting increased fluency. In contrast, interactions with a child-voiced interface led to fewer, but more elaborated responses, indicating more elaboration. These results suggest that the perceived age of AI voice in CAs can have an effect on creativity expression. Adult voices may be better suited for encouraging idea generation, while child voices may support imaginative depth and rich elaboration.

According to our initial hypothesis grounded in interdisciplinary theory basis, we anticipated that children might respond differently to adult versus child-voice interfaces, with the child-like voice interface potentially affording more elaborative expression due to more agency children perceive in peer relationships in relation to relationships with adults, whereas adult-voice interface would encourage children to produce more ideas in effective and systematic way with less elaboration. However, we did not initially predict that similar effects could also emerge among adult participants.

While the study offers valuable insights into how AI voice characteristics in CAs affect creativity expression, there are following limitations must be acknowledged. First, a group confound effect may have influenced the results, Each group of participants was exposed to only one voice type (either adult or child), preventing individual level comparisons across conditions. Additionally, the small sample size and potential variation in participants previous exposure to similar tasks could have influenced the results. The subjective nature of interpreting creative responses, even with structured analysis and guides, introduces some degree of bias. Furthermore, the individual expression style of participants was not controlled in this study setting.

The idea to better the future study settings conducted with children could include a pre-testing of children's creativity expression style, and assigning children into study groups quasi-randomly balancing each group according to children's age, gender and pre-test creativity score.

4.3 Future Study Settings

The project continues in June 2025 with children participating in the study. In the scope of this Master's thesis we have analysed the voice interfaces effects of participants fluency and elaboration, and the next phase of the project will also include the analysis of participants originality scores, in both adult and child participants, and seeing if and how they correlate to perceived age of the AI voice in CAs.

To be able to avoid the risk of confound bias in small sample sizes based on participants individual creativity expression styles, we have created a design for pre-testing participants creative expression style to ensure more balanced study groups to be used in the future study settings.

4.3.1 Creativity Pre-Test

The goal of the creativity pre-test is not to make assumptions about the children's overall creative abilities, but rather to serve as a control for individual differences in styles of creative expression. This is particularly important in studies with smaller sample sizes, where unbalanced distributions of creative expression could introduce bias into the results. The pre-test is based on The TTCT (Torrance Tests of Creative Thinking) that originally has two versions: the TTCT-Verbal and the TTCT-Figural [61]. Our pre-test is based on The TTCT-Verbal which consists of five different activities designed to assess verbal creativity in following three dimensions: fluency, flexibility and originality. [61]. These activities are: "ask-and-guess," "product improvement," "unusual uses," "unusual questions," and "just suppose." [117] To assess verbal creativity with TTCT-Verbal, minimun three categories of activities should be used [61].

For the purpose of our pre-test to be conducted on children before they participate in the actual study we have chosen following three verbal tasks, that are modified to correspond to the theme of robots and other CAs:

The tasks are the following:

1. **Unusual Uses:** Children are asked to think of as many uses they can for a known items. *„List as many uses as possible for a robot"*
2. **Ask-and-Guess:** Children are asked to guess causes and consequences, which requires them to think about situations from different perspectives. *„What would happen if every child had a robot?"*
3. **Product Improvement:** Children are asked to think of ways to change or improve a toy or an object, encouraging them to think about how things work and how they might be made to work better. *„Imagine you had your own robot. What would you want it be able to do?"*

Audio recording will be used to record children's answer, giving every child same opportunity to express themselves without having to write the answers down. Answers will be transcribed and categorized for analysis. The Verbal TTCT pre-test will be conducted to children as an interview by one of the researchers. Interviewer asks the children to generate as many ideas as they can to these three questions, giving them 4 minutes per question (12 minutes total).

4.3.2 Rubric Scoring System of The Pre-Test

The rubric scoring of TTCT that we have applied to in our study setting follows the scoring system of the adapted version of the TTCT [118], specifically the Verbal-A form of the test [61]. Rubric scoring of TTCT is a structured and consistent evaluation method that is based on a set of defined criteria for each component [118]. This rubric scoring system enables the production of a combined creativity score that integrates multiple dimensions of creative thinking. L. Rababah's adaptation of the TTCT rubric was specifically modified for practical and easy use in school settings in 2018, enabling teachers to more easily identify highly creative students who might be overlooked by other assessments. Their version scores creativity across three dimensions, and sets threshold values to categorise individuals' creativity levels.

Scoring in the rubric scoring system happens in the following way, counting the rubric score of each participants answers to each of the three tasks separately.

Fluency (max. 12 points per task)

Measuring happens manually dissecting the different ideas the participant expresses to each of the tasks separately so that each idea that the participants presented to the task can score 0-3 points according the following criteria:

- 0 points - no answer
- 1 point - Unrelated or repeated idea
- 2 points - Incomplete, partial or underdeveloped response (*"a hat"*, *"a person sitting"*)
- 3 points - Complete, fully-formed idea that is well explained (*"it's a giant cat that sits on a table and looks at the moon"*)

Flexibility (max. 12 points per task)

Flexibility is equivalent to the concept of elaboration, defined as an ability to produce a variety of ideas, shift from one approach to another, or use a variety of strategies (various categories in answers). Flexibility is about how many different categories or themes a person addresses in their responses. Measuring flexibility will be done by manually dissecting the data and assigning the scores according to the degree of diversity of each idea presented. Each unique idea presented by participants will score 0 to 3 points, points adding up to maximum of 12 points per task by following criteria:

- 0 points - no answer
- 1 point - No transformation or development of categories (*"a hat"*, *"a big, round table"*)
- 2 points - Incomplete transformation or development of categories (*"a dog that sits on the table"*)
- 3 points - Complete transformation or development of categories (*"a giant cat that sits on a table and looks at the moon"*)

Originality (max. 12 points per task)

Originality, the ability to produce ideas that are unusual and unique, is counted by giving each each idea presented an originality scores between 0 to 3 points, reaching up to maximum of 12 points per task. Measuring originality happens by manually comparing participants answer to the existing database of answers, in this case, comparing participants answer to all the other answers children participating in the study have given.

- 0 points - no answer
- 1 point - No originality in answer
- 2 points - Originality that appears rarely
- 3 points - Originality that appears more frequently

Creativity Score

Each task completed by a participant results a score ranging from 0 to 36 points. After summing the scores across all tasks, a total score is calculated for each participant. The average of these totals represents the participant's overall creativity score. Based on the distribution of these scores, a cutoff point can be selected to divide participants into two groups: those with higher creative expression and those with lower. This division allows for the formation of study groups with a balanced number of high- and low-scoring participants, helping to control for potential biases in creativity level across experimental conditions.

Chapter 5

Discussion

5.1 The Historical and Structural Biases of HRI Research

The development of HRI research, due to historical and structural reasons, has been solely technically driven by engineering-focus. K. Richardson, British anthropologist and ethicist best known for their work on the ethics of robotics and artificial intelligence, illustrates this dynamic in their ethnographic work at the MIT lab, noting that violence and robotics are often "frequent bedfellows"[119]. Although the lab was officially part of the Artificial Intelligence Laboratory, it physically shared space with MIT's Computer Science department, and when MIT researchers were working on creating a "robot child," the project was funded by a grant from *the Defence Advanced Research Projects Agency*, DARPA [119]. This example highlights how deeply embedded robotics R&D is within broader systems of power, particularly military and technological institutions, which inevitably shape the kinds of robots that are imagined and built.

J. M. Jordan, an academic and technology analyst who writes about emerging technologies and their implications for business and society, has noted that once these core functions are established, attention turns to potential real-world applications, but often without considering the social contexts regarding application and end users [120]. This highlights that engineering and technical expertise form the core of robotics and are not designed to address broader social and ethical concerns. This approach echoes the early development of computers, where usability and user experience were secondary concerns for a long time but there is now a good amount of research dedicated designing interfaces with their users in mind, (e.g. Technology Acceptance Model (TAM) and User Experience (UX) design models) [121].

A significant structural challenge, as we came to face in technically oriented academic research lies in the uncritical adoption of certain research methods and metrics as standardised benchmarks. Once a method is published in a peer-reviewed journal, it gains a form of legitimacy that encourages its replication across diverse studies, not necessarily because it is the best conceptual fit, but because it is considered reliable and reproducible. This tendency often sidelines critical reflection on whether these methods truly capture the complex, context-dependent social and psychological phenomena under investigation, such as creativity.

Our own experience illustrates this issue. When reviewing prior research on how creativity in children has been measured within the context of HRI research, we found many of the existing approaches difficult to reconcile with the aims of our study, which focuses on children's creative responses to different artificially generated voices. Expressing our interest in exploring this topic with a study design that is not benchmarked provoked mixed reactions. Colleagues from engineering focised background voiced concerns over the challenges of measurement, while our partners at The LEGO Group responded with enthusiasm, appreciating the novelty of our inquiry ("I don't think I've seen research like that before").

The broader technically oriented research community frequently overlooks how deeply context-sensitive constructs like creativity are. The transfer of measurement tools across settings without revisiting their conceptual grounding leads to their widespread but potentially misplaced use. This problem is compounded by the fact that peer reviewers, often rooted in technically driven disciplines, tend to prioritize reproducibility and quantifiability over conceptual nuance. Consequently, methodological limitations tied to the social complexity of HRI research can remain unchallenged. Over time, these methods become entrenched as benchmarks, even when they offer superficial or reductive interpretations of rich, multifaceted social phenomena. [122]

This dynamic creates a structural incentive for researchers to "play it safe" by employing established measures, even when these may be conceptually insufficient or theoretically weak. More context-sensitive or socially informed methodologies are often perceived as risky, particularly when researchers face pressure to produce publishable results quickly. This discourages methodological innovation and reinforces a cycle in which standardized, yet potentially flawed, metrics dominate.

To advance the field, it is necessary that researchers ground their studies in clearly articulated theoretical frameworks that are specific to their unique research context. This entails explicitly defining the core concept in relation to the phenomena being examined, carefully selecting the dimensions most relevant to the study's aims, and transparently justifying the choice of measurement tools. Only through such deliberate and reflective practice can research move beyond superficial assessments toward a richer, more accurate understanding of creativity and other complex human-centred constructs.

This thesis has shown that quantitative data do not "speak for themselves", instead it is always interpreted through theoretical frameworks shaped by methodological choices and institutional norms. The process of moving from data to conclusion is neither automatic nor value-neutral but it involves interpretative labour and always reflects the epistemological assumptions of the researchers. In the context HRI where the aim is to understand complex socio-technical phenomena, it highlights the necessity of interdisciplinary research particularly when addressing abstract and context-dependent concepts such as creativity.

5.2 Missing Child Perspective

As we showed in the Methodology chapter, the gap in knowledge and focus becomes even more pronounced when it comes to children. Despite the growing presence of research on robot and children's interaction in education, therapy, and play, we could very rarely draw learnings on voice preferences from those studies due to the lack of comparisonn or even acknowledgement of the voice as a possible variable. If we are to design meaningful and ethical robotic products for children, we must examine a wide range of factors from developmental needs and emotional responses to cultural and ethical implications, as well as the high standards we have on technical performance and human safety.

R House, a Human-Robot Interaction (HRI) lab in Indiana, conducted interviews with 100 robotic researchers worldwide over a nine-year period, yet only one researcher mentioned children in their interview summary [123]. This highlights a gap in the current research: despite the references to children in anthropological and technological discussions [119], [124]–[126], the deeper implications of children's interactions with robots are underexplored. Much of the existing anthropological literature on children and robots tends to focus on how children interact with robots in ways similar to their interactions with other humans. These studies often explore how robots are designed to simulate a "caregiver-child relationship" [119, p. 61] or how children's acceptance of robots can reveal the boundaries of what is considered human-like [126, p. 146]. However, these discussions tend to remain limited to behavioural observations, without exploring deeper into the developmental and societal implications of such interactions.

Also in Danish context there are already a number of studies about people interacting with humanoid robots over time [127]–[131], but children-focused studies are more limited.

Children are capable of engaging in rich and nuanced interactions with robots, often perceiving them as social others, entities that are treated as socially responsive and emotionally present [132]–[134]. Children's interaction with robots is also highly context dependent, as it is demonstrated in the earlier research that children's acceptance and perception of robots as social companions is influenced both by their previous experiences and cultural narratives around technology, personhood, and animacy [135], [136].

A growing body of research shows that robots can be especially effective as supportive agents in therapeutic and educational settings, frequently outperforming humans or screens, particularly for children with autism spectrum disorder (ASD) [137]. One explanation is that robots reduce the cognitive and emotional load of social interaction while still eliciting meaningful engagement. Importantly, there is evidence that these interactions can help train social skills that transfer to human contexts [138]. This underlines the need for more child-specific research on voice preferences in HRI to understand how, why, and under what conditions robots can best support children's development.

5.3 Connecting Metrics, Meaning and Machines

Finding ways to balance the objectivity of metrics and keeping contextualization of those metrics in mind was both a challenging and enriching experience. It required negotiation of values, priorities, and methodological assumptions. We learned to bring together the strengths of both approaches and find a way to build a quantitative study around ways to support creative expression through play. The project we developed reflects this synthesis and has received positive feedback from both institutions. It has also generated significant interest across different departments at the LEGO Group, highlighting the value of cross-disciplinary thinking in the future of child-AI interaction research.

As Techno-Anthropology master's students, we have used this collaboration opportunity to navigate and reflect on the distinct approaches that these different institutions have toward product development, system design, and research. Throughout this process, we have learned to communicate across disciplinary boundaries, translating between the languages and values of engineering, design, and social science.

In this thesis, we are using the learnings and experiences gained through the project to critically examine the gap between standardised, metric-focused research and anthropological approaches that emphasise contextual, relational, and ethical dimensions of technology use. We argue that in order to design socially meaningful and effective interactive systems, particularly for children, there is a need to integrate these perspectives more deliberately. Our aim is not only to conceptualise this need at a macro level by looking at broader systemic and cultural influences, but also to examine how micro-level social dynamics, such as children's lived experiences, can affect research outcomes.

To ensure that these frameworks and guidelines are well grounded, theoretical and empirical research on the impacts of these technologies is essential. We need critical studies that address not only the functionality of these systems but also their broader societal implications. Conversational robots and voice-enabled AI systems represent a new kind of social relationship between humans and technology. Without research grounded in social theory and contextual awareness, there is a real danger that these technologies will be introduced into children's lives without adequate understanding of their long-term effects. In such cases, AI systems risk not only falling short of supporting positive growth but also may unintentionally contribute to harmful or undesirable outcomes for young users.

The aim of this thesis is not to pass judgment on whether or not children should have access to emerging AI technology, like conversational robots, because even if we tried, we believe there is not enough information to make such an absolute decision. We focus on reflecting the current knowledge production about conversational systems in empirical research and see if the research theorises about the implications these technologies might have on children and their rights, and hopefully add to that knowledge making in a meaningful way.

There are many different ways to produce knowledge, and we have focused on academic knowledge production in this thesis. Specifically analysing sociology of knowledge through *Sociology of Scientific Knowledge* (SSK) to reflect on qualitative and quantitative methods as the ways we saw knowledge being produced in our internships.

5.4 Limitations and reflections

There are critical challenges to be considered while researching children and AI interactions, particularly vulnerable populations such as children. The social, emotional, and developmental implications of such technologies are not yet fully known.

We experienced many bumps in the road when designing our research study that became valuable learning experiences. In the beginning when we were exploring the possibility of training a small children-specific AI to allow for a guided play scenario in our research, we discovered fundamental technical problems that made it clear to us that we could not do research with children interacting with a AI, at least not within the timeframe of this thesis:

- **Lack of datasets on childrens culture**. We found very minimal public datasets with *childrens-specific culture* to train on. We searched for datasets on child TV shows, songs or anything else child-specific and found nothing ready-made that we could use in the short amount of time we had for the research.

- **Lack of datasets on childrens voices**. There are now numerous websites where you can get access to a selection of childrens voices through a Text-to-speech synthesiser. But public datasets to train an AI on were not available, or the resources we reached out to did not respond to our request for access within the needed time limit.

- **Child speech is hard for AI to recognize**. AI speech recognition has not been trained on childrens voices, so their speech, pauses and vocal gestures etc. can be hard for speech recognition software to understand [33].

In all cases, the underlying problem is a lack of available training data. If we would've had access to adequate training data, we could've explored the ethical considerations further of having children interact with AI in a research setting. But with the resources and the datasets we had access to, it's unlikely that it would've made it through the ethical committee in time. Especially because we don't even know if we'd be able to ensure that the AI would understand children's voices, respond in an age-appropriate way or stay aligned to the game setting.

Additionally, we acknowledge that there are inherent risks in designing AI for children without grounding in social theory or developmental psychology. Quantifying children's social behavior without considering context can result in superficial findings—or worse, lead to designs that mislead or harm users. Systems that ignore children's emotional and cognitive development may produce interactions that are irrelevant, confusing, or manipulative. Designing conversational AI systems without sufficient sensitivity to children's social and developmental needs risks producing interactions that are inappropriate or disconnected from real-world. Since children are already high adopters of AI generative technologies, this highlights the urgent need for more robust interdisciplinary approaches that integrate social theory, ethics, and contextual understanding into the technical design and evaluation of these systems.

For our findings, we critically examined ourselves throughout the process, not only the tools we use to measure creativity, but also the assumptions and values embedded within them. We believe, to overcome these validity challenges, research on creatvity must:

- Make explicit the theoretical foundations of creativity metrics, and grounding them in research
- Acknowledge and account for the cultural, linguistic, and social positionalities of evaluators
- Provide transparency in coding systems and justification for why these measurement tools were used

It was inspiring to see how qualitative research methods were used in The LEGO Group's approach to studying children's experiences. Their framework was iterative, reflective, and attentive to the perspectives of the children involved. Ethical considerations were not treated as separate from the research process, but as an integral part of designing meaningful and respectful interactions with young participants. AAU Robotics and Automation puts a strong emphasis on systematic, quantitative research methods, prioritising replicability, scalability, and technical precision, which are crucial for advancing engineering-led development of robotic systems. We found both institutions were supportive and interested in helping us in every way they could, and mixing these methods was challenging, but an extremely valuable experience.

5.5 Towards Future Approaches

This has important implications for research contexts where creativity is measured or evaluated, especially in interdisciplinary or applied settings such as education, design, or human-technology interaction. Decisions about how creativity is defined and operationalised are not just technical or methodological choices but also socio-cultural ones. These decisions carry embedded assumptions about what kinds of behaviours or outputs are valued and how they should be measured. As such, it is crucial that researchers approach these decisions critically and reflexively, considering the broader disciplinary and societal frameworks in which their work is situated.

The UNICEF published a report about children's rights in 2012, recognizing children as one of the most marginalised and vulnerable groups in society, often excluded from decision-making processes that directly impact them, and stresses the importance of including children's voices in the development of systems that affect them [139]. Given the vulnerabilities associated with robots for children and the under-researched nature of this field, there is a pressing need for a more interdisciplinary approach to robotics R&D. Incorporating insights from social sciences can help ensure that robots designed for children are ethical and developmentally appropriate.

Also by shifting perspectives and involving children in the design process, we can open up new opportunities to create robots that are not only fun, creative, and diverse but also free from the biases and social narratives shaped by adult perspectives. This approach would ensure that robots better align with children's needs, and promote positive developmental outcomes.

But if we shift the focus and think of robots as tools that afford humans with experiences and train valuable skills, we could start training non-native English speaking children through games, because children are more willing to speak, and make mistakes, when presented with Keepon, a non-speaking robot [140]. Toy robots have up to this point mostly been unilateral, non-verbal interactions, or verbal interactions that with a robotic, slow and non-interactive voice. The technology of artificially creating human-like voices with the conversational capabilities of artificial agents (Chatbots) is a subject that needs to be studied further in social interactions of "the real world". We know that children connect to robots on a deep level, and we know that they don't need to be humanoid in appearance for children to connect to them, but we don't know how combining a chatbot into a robot will affect their relationships to this object.

A valuable direction for future research would be a comprehensive critical analysis of creativity-related studies within the fields of human-robot interaction and technology. This analysis should focus specifically on the methods used to measure creativity, systematically examining how creativity is operationalized across different studies. Key areas of attention would include limitations such as context insensitivity, theoretical misalignment, and the inappropriate use of creativity assessment tools. Such a review would offer important insights into the challenges of evaluating creativity in technology-mediated settings and could guide the development of more contextually responsive and theoretically grounded measurement approaches.

Maria Sylvia Alaoja
malaoj23@student.aau.dk
Guðfinna Kristinsdóttir
gkrist23@student.aau.dk

Bibliography

[1] UNICEF and LEGO Foundation, *Learning through play: Strengthening learning through play in early childhood education programmes*, `https://www.unicef.org/sites/default/files/2018-12/UNICEF-Lego-Foundation-Learning-through-Play.pdf`, 2018.

[2] T. H. Eriksen, *Small places, large issues: An introduction to social and cultural anthropology*. PLUTO press, 2015.

[3] M. Venckutė, I. Mulvik, B. Lucas, and P. Kampylis, "Creativity—a transversal skill for lifelong learning," *An Overview of Existing Concepts and Practices: Final Report*, 2020.

[4] P. Y. Mullineaux and L. F. Dilalla, "Preschool pretend play behaviors and early adolescent creativity," *The Journal of Creative Behavior*, vol. 43, no. 1, pp. 41–57, 2009.

[5] K. Mannheim, *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. Chicago: University of Chicago Press, 1936.

[6] H. Collins, "Scientific knowledge, sociology of," in *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, J. D. Wright, Ed., Amsterdam: Elsevier, 2015, pp. 139–145.

[7] S. Shapin, "Here and everywhere: Sociology of scientific knowledge," *Annual review of sociology*, vol. 21, no. 1, pp. 289–321, 1995.

[8] S. Yearley, "Understanding science from the perspective of the sociology of scientific knowledge: An overview," *Public Understanding of Science*, vol. 3, no. 3, pp. 245–258, Jul. 1994.

[9] A. Toosi, A. G. Bottino, B. Saboury, E. Siegel, and A. Rahmim, "A brief history of ai: How to prevent another winter (a critical review)," *PET clinics*, vol. 16, no. 4, pp. 449–469, 2021.

[10] B. G. Buchanan, "A (very) brief history of artificial intelligence," *Ai Magazine*, vol. 26, no. 4, pp. 53–53, 2005.

[11] K. Li, "Shaw and automata," in *Bernard Shaw, Automata, Robots, and Artificial Intelligence*. Cham: Springer Nature Switzerland, 2024, pp. 9–26.

[12] D. J. de Solla Price, "Automata and the origins of mechanism and mechanistic philosophy," *Technology and Culture*, vol. 5, no. 1, pp. 9–23, 1964.

[13] E. S. and, "The mechanical turk: A short history of 'artificial artificial intelligence'," *Cultural Studies*, vol. 37, no. 1, pp. 65–87, 2023. eprint: `https://doi.org/10.1080/09502386.2022.2042580`.

[14] F.-F. Li, *The Worlds I See: Curiosity, exploration, and discovery at the dawn of AI*. Flatiron books: a moment of lift book, 2023.

[15] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," *Journal of Industrial Information Integration*, vol. 23, p. 100 224, Sep. 1, 2021.
[16] C. Gordon. "ChatGPT is the fastest growing app in the history of web applications," Forbes. (Feb. 3, 2023), [Online]. Available: `https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/` (visited on 11/20/2024).
[17] A. Bick, A. Blandin, and D. Deming, *The rapid adoption of generative ai | cepr*, Oct. 2024.
[18] L. Yee, M. Chui, and R. Roberts, *Why agents are the next frontier of generative ai*, Jul. 2024.
[19] I. Belcic and C. Stryker, *Ai agents in 2025: Expectations vs. reality*, Apr. 2025.
[20] H. Srivatsa, *Beyond the beep: Why voice ai agents are the next strategic imperative for ai product leaders*, Apr. 2025.
[21] S. Altman and J. Ive, *Sam and jony introduce io*, Mar. 2025.
[22] *Project astra | exploring the future of learning with an ai tutor research prototype*, Google DeepMind, 2025.
[23] G. V. Mar Negreiro, "Children and generative ai," OECD, Paris, 1, 2025.
[24] OECD, "Pisa 2022 results (volume i): The state of learning and equity in education," OECD, Paris, 1, 2023.
[25] D. Hendrycks, M. Mazeika, and T. Woodside, *Ai risks that could lead to catastrophe: Cais*, Oct. 2023.
[26] OECD.
[27] E. Commission, *Artificial intelligence – questions and answers**.
[28] I. Department for Science and Technology, *Safety and security risks of generative artificial intelligence to 2025 (annex b)*.
[29] N. I. of Standards and Technology, *Ai risks and trustworthiness*, Mar. 2025.
[30] N. S. Center, *Ai and cyber security: What you need to know*.
[31] A. for Digital Government, *The danish national strategy for artificial intelligence*.
[32] V. Andries and J. Robertson, "Alexa doesn't have that many feelings: Children's understanding of ai through interactions with smart speakers in their homes," *Computers and Education: Artificial Intelligence*, vol. 5, p. 100 176, 2023.
[33] Y. Xu, S. Branham, X. Deng, P. Collins, and M. Warschauer, "Are current voice interfaces designed to support children's language development?" In *Proceedings of the 2021 CHI conference on human factors in computing systems*, 2021, pp. 1–12.
[34] R. Garg, H. Cui, S. Seligson, *et al.*, "The last decade of hci research on children and voice-based conversational agents," in *Proceedings of the 2022 CHI conference on human factors in computing systems*, 2022, pp. 1–19.
[35] D. F. Lancy, *The Anthropology of Childhood: Cherubs, Chattel, Changelings*. Cambridge University Press, Mar. 10, 2022.
[36] L. Cai, "Realities vs expectations: Children's perception and imagination of ai," *International Journal of Technology and Design Education*, vol. 34, no. 5, pp. 1687–1704, 2024.
[37] E. Kosoy, E. R. Reagan, and S. Jeong, "Adventures in ai wonderland: How children are shaping the future of ai," *Interactions*, vol. 31, no. 6, pp. 28–33, 2024.

[38] M. Newman, K. Sun, I. B. Dalla Gasperina, *et al.*, "" i want it to talk like darth vader": Helping children construct creative self-efficacy with generative ai," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–18.

[39] M. P. Aylett, S. J. Sutton, and Y. Vazquez-Alvarez, "The right kind of unnatural: Designing a robot voice," in *Proceedings of the 1st international conference on conversational user interfaces*, 2019, pp. 1–2.

[40] Y. Xu and M. Warschauer, "What are you talking to?: Understanding children's perceptions of conversational agents," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–13.

[41] M. Scaife and M. van Duuren, "Do computers have brains? what children believe about intelligent artifacts," *British Journal of Developmental Psychology*, vol. 13, no. 4, pp. 367–377, 1995.

[42] M. van Duuren and M. Scaife, ""because a robot's brain hasn't got a brain, it just controls itself"—children's attributions of brain related behaviour to intelligent artefacts," *European Journal of Psychology of Education*, vol. 11, pp. 365–376, 1996.

[43] L. Hubbard, S. Ding, V. Le, P. Kim, and T. Yeh, "Voice design to support young children's agency in child-agent interaction," in *Proceedings of the 3rd Conference on Conversational User Interfaces*, 2021, pp. 1–10.

[44] S. Druga, R. Williams, C. Breazeal, and M. Resnick, "" hey google is it ok if i eat you?" initial explorations in child-agent interaction," in *Proceedings of the 2017 conference on interaction design and children*, 2017, pp. 595–600.

[45] Y. Yuan, S. Thompson, K. Watson, *et al.*, "Speech interface reformulations and voice assistant personification preferences of children and parents," *International Journal of Child-Computer Interaction*, vol. 21, pp. 77–88, 2019.

[46] G. Wang, J. Zhao, M. Van Kleek, and N. Shadbolt, "Challenges and opportunities in translating ethical ai principles into practice for children," *Nature Machine Intelligence*, vol. 6, no. 3, pp. 265–270, 2024.

[47] Y. Yu, Y. Liu, J. Zhang, Y. Huang, and Y. Wang, *Understanding generative ai risks for youth: A taxonomy based on empirical data*, 2025. arXiv: `2502.16383 [cs.HC]`.

[48] I. R. Berson, M. J. Berson, and W. Luo, "Innovating responsibly: Ethical considerations for ai in early childhood education," *AI, Brain and Child*, vol. 1, no. 1, p. 2, 2025.

[49] M. Cuciniello, T. Amorese, G. Cordasco, *et al.*, *Identifying synthetic voices qualities for conversational agents*, 2022. arXiv: `2205.04149 [cs.HC]`.

[50] S. B. H. Pias, R. Huang, D. S. Williamson, M. Kim, and A. Kapadia, "The impact of perceived tone, age, and gender on voice assistant persuasiveness in the context of product recommendations," in *ACM Conversational User Interfaces 2024*, ser. CUI '24, ACM, Jul. 2024, pp. 1–15.

[51] G. Zellou, M. Cohn, and B. Ferenc Segedin, "Age-and gender-related differences in speech alignment toward humans and voice-ai," *Frontiers in Communication*, vol. 5, p. 600 361, 2021.

[52] T. L. Foundation, "What we mean by: Learning through play," 1.2, 2017.

[53] J. N. Zosh, E. J. Hopkins, H. Jensen, *et al.*, *Learning through Play: A Review of the Evidence*. Billund, Denmark: LEGO Foundation, 2017.

[54] P. H. Kahn, T. Kanda, H. Ishiguro, *et al.*, "Human creativity can be facilitated through interacting with a social robot," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 173–180.

[55] S. Ali, N. Devasia, H. W. Park, and C. Breazeal, “Social robots as creativity eliciting agents,” *Frontiers in Robotics and AI*, vol. 8, p. 673 730, 2021.
[56] T. Cadierno, M. B. Hansen, J. T. Lauridsen, S. W. Eskildsen, K. Fenyvesi, and S. Hannibal Jensen, “A multi-factor study of the development of english receptive skills by young danish children,” *Bilingualism: Language and Cognition*, vol. 25, no. 4, pp. 566–578, 2022.
[57] T. M. Amabile, “The social psychology of creativity: A componential conceptualization,” *Journal of Personality and Social Psychology*, vol. 45, no. 2, pp. 357–376, 1983.
[58] R. J. Sternberg and T. I. Lubart, *Defying the Crowd: Cultivating Creativity in a Culture of Conformity*. New York: The Free Press, 1995.
[59] E. P. Torrance, “A longitudinal examination of the fourth grade slump in creativity,” *Gifted Child Quarterly*, vol. 12, no. 4, pp. 195–199, 1968.
[60] UNICEF, *Policy guidance on ai for children*, 2021.
[61] E. P. Torrance, *The Torrance Tests of Creative Thinking - Interpretive Manual. A & B*. Bensenville, IL: Scholastic Testing Service, Inc., 2018.
[62] S. S. Metwaly, W. V. den Noortgate, and E. Kyndt, “Approaches to measuring creativity: A systematic literature review,” *Creativity Theories - Research - Applications*, vol. 4, no. 2, pp. 1–12, 2017.
[63] E. P. Torrance, *Torrance® Tests of Creative Thinking: Interpretive Manual*. Scholastic Testing Service, 2018.
[64] B. Barbot, M. Besançon, and T. I. Lubart, “Creative potential in educational settings: Its nature, measure, and nurture,” *Education 3-13*, vol. 43, pp. 371–381, 2015.
[65] R. MacDonald, D. Miell, and L. Morgan, “Social processes and creative collaboration in children,” *European Journal of Psychology of Education*, vol. 15, no. 4, pp. 405–415, 2000.
[66] M. Baas, C. K. W. De Dreu, and B. A. Nijstad, “A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus?” *Psychological Bulletin*, vol. 134, no. 6, pp. 779–806, 2008.
[67] D. Miell and R. MacDonald, “Children’s creative collaborations: The importance of friendship when working together on a musical composition,” *Social Development*, vol. 9, no. 3, pp. 348–369, 2000.
[68] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, Massachusetts: Harvard University Press, 1978.
[69] S. Gürdal and E. Sorbring, “Children’s agency in parent–child, teacher–pupil and peer relationship contexts,” *International Journal of Qualitative Studies on Health and Well-Being*, vol. 13, no. sup1, 2018.
[70] E. Fosch-Villaronga, S. van der Hof, C. Lutz, and A. Tamò-Larrieux, “Toy story or children story? putting children and their rights at the forefront of the artificial intelligence revolution,” *AI & SOCIETY*, vol. 38, no. 1, pp. 133–152, Feb. 1, 2023.
[71] P. H. Kahn Jr, T. Kanda, H. Ishiguro, S. Shen, H. E. Gary, and J. H. Ruckert, “Creative collaboration with a social robot,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 2014, pp. 99–103.
[72] P. Alves-Oliveira, S. Tulli, P. Wilken, R. Merhej, J. Gandum, and A. Paiva, “Sparking creativity with robots: A design perspective,” in *14th annual ACM/IEEE international conference on human robot interaction (HRI)*, IEEE, 2019.

[73] P. Alves-Oliveira, P. Arriaga, M. A. Cronin, and A. Paiva, "Creativity encounters between children and robots," in *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, 2020, pp. 379–388.
[74] C. Oertel *et al.*, "Engagement in human-robot interactions: An overview," *Frontiers in Robotics and AI*, vol. 7, 2020.
[75] S. P. Besemer and D. J. Treffinger, "Analysis of creative products: Review and synthesis," *The Journal of Creative Behavior*, vol. 15, no. 3, pp. 158–178, 1981.
[76] D. Hocevar, "Measurement of creativity: Review and critique," *Journal of Personality Assessment*, vol. 45, no. 5, pp. 450–464, 1981.
[77] B. Barbot, M. Besançon, and T. I. Lubart, "Assessing creativity in the classroom," *The Open Education Journal*, vol. 4, pp. 124–132, 2011.
[78] D. J. Treffinger, *Creativity, creative thinking, and critical thinking: In search of definitions*. Sarasota, FL: Center for Creative Learning, 1996.
[79] R. P. Weiner, *Creativity and Beyond: Cultures, Values, and Change*. Albany, NY: State University of New York Press, 2012.
[80] T. Ivancovsky, O. Kleinmintz, J. Lee, J. Kurman, and S. G. Shamay-Tsoory, "The neural underpinnings of cross-cultural differences in creativity," *Human Brain Mapping*, 2018.
[81] R. Mehta and D. W. Dahl, "Creativity: Past, present, and future," *Consumer Psychology Review*, vol. 2, no. 1, pp. 30–49, 2019.
[82] M. W. Eysenck and M. T. Keane, *Cognitive Psychology: A Student's Handbook*, 5th. New York: Psychology Press, 2005.
[83] G. A. Miller, "The cognitive revolution: A historical perspective," *Trends in Cognitive Sciences*, vol. 7, no. 3, pp. 141–144, 2003.
[84] A. Kozbelt and Y. Durmysheva, "Lifespan creativity in a non-western artistic tradition: A study of japanese ukiyo-e printmakers," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 65, no. 1, 2023.
[85] A. Newell, J. C. Shaw, and H. A. Simon, "The processes of creative thinking," in *Contemporary Approaches to Creative Thinking*, H. E. Gruber, G. Terrell, and M. Wertheimer, Eds., New York: Atherton Press, 1962, pp. 63–119.
[86] M. Dellas and E. L. Gaier, "Identification of creativity: The individual," *Psychological Bulletin*, vol. 73, no. 1, pp. 55–73, 1970.
[87] R. W. Skager, C. B. Schultz, and S. P. Klein, "Quality and quantity of accomplishments as measures of creativity," *Journal of Educational Psychology*, vol. 56, no. 1, pp. 31–39, 1965.
[88] V. Hänninen, J. Partanen, and O.-H. Ylijoki, *Sosiaalipsykologian suunnannäyttäjiä*, Finnish, 3. painos. Tampere: Vastapaino, 2017.
[89] S. S. Metwaly, W. V. den Noortgate, and E. Kyndt, "Approaches to measuring creativity: A systematic literature review," *Creativity Theories – Research – Applications*, vol. 4, no. 2, pp. 1–17, 2017.
[90] R. J. Sternberg and T. I. Lubart, "The concept of creativity: Prospects and paradigms," in *Handbook of Creativity*, Cambridge University Press, 1999, pp. 3–15.
[91] M. A. Runco, *Divergent Thinking*. Norwood, NJ: Ablex, 1991.
[92] J. Peter H. Kahn, B. Friedman, R. L. Severson, and E. N. Feldman, "Creativity tasks and coding system – used in the plasma display window study," University of Washington, UW Information School, Seattle, WA, Tech. Rep. IS-TR-2005-04-01, Apr. 2005.

[93] W. Niu and R. Sternberg, "Contemporary studies on the concept of creativity: The east and the west," *The Journal of Creative Behavior*, vol. 36, no. 4, pp. 269–288, 2002.
[94] H. C. Kim and H. R. Markus, "Deviance or uniqueness, harmony or conformity: A cultural analysis," *Journal of Personality and Social Psychology*, vol. 77, no. 4, pp. 785–800, 1999.
[95] H. H. Kim, S. Mishra, P. Hinds, and L. Liu, "Creativity and culture: State of the art," *Design thinking research: Studying co-creation in practice*, pp. 75–85, 2011.
[96] K. H. Kim and D. L. Zabelina, "Cultural bias in assessment: Can creativity assessment help?" *International Journal of Critical Pedagogy*, vol. 6, no. 2, 2015.
[97] M. Erez and R. Nouri, "Creativity: The influence of cultural, social, and work contexts," *Management and Organization Review*, vol. 6, no. 3, pp. 351–370, 2010.
[98] D. Dumas and K. N. Dunbar, "Understanding fluency and originality: A latent variable perspective," *Thinking Skills and Creativity*, vol. 14, pp. 56–67, 2014.
[99] M. Runco and W. Mraz, "Scoring divergent thinking tests using total ideational output and a creativity index," *Educational and Psychological Measurement*, vol. 52, no. 1, pp. 213–221, 1992.
[100] D. H. Paulus and J. S. Renzuli, "Scoring creativity tests by computer," *Gifted Child Quarterly*, vol. 12, no. 2, pp. 79–84, 1968.
[101] J. C. Kaufman, J. Lee, J. Baer, and S. Lee, "Captions, consistency, creativity, and the consensual assessment technique: New evidence of reliability," *Thinking Skills and Creativity*, vol. 2, no. 2, pp. 96–106, 2007.
[102] M. A. Runco and S. Acar, "Divergent thinking as an indicator of creative potential," *Creativity Research Journal*, vol. 24, no. 1, pp. 66–75, 2012.
[103] E. Rákóczi and I. Szitó, "Using a creativity test standardized on a hungarian sample to identify high creative abilities of socially disadvantaged students," *South Florida Journal of Development*, vol. 2, no. 2, pp. 2427–2437, Apr. 2021.
[104] M. A. Runco and S. Acar, "Do tests of divergent thinking have an experiential bias?" *Psychology of Aesthetics, Creativity, and the Arts*, vol. 4, no. 3, pp. 144–148, 2010.
[105] A. M. A. Alabbasi, S. H. Paek, D. Kim, and B. Cramond, "What do educators need to know about the torrance tests of creative thinking: A comprehensive review," *Frontiers in Psychology*, vol. 13, Oct. 2022.
[106] A. O. Hughes and J. S. Drew, "A state creative?" *Papers in the Social Sciences*, 1984.
[107] M. Blamires and A. Peterson, "Can creativity be assessed? towards an evidence informed framework for assessing and planning progress in creativity," *Cambridge Journal of Education*, vol. 44, no. 2, pp. 147–162, 2014.
[108] E. P. Torrance, "The rationale and validation of the torrance tests of creative thinking, verbal and figural," *Tempo*, vol. 20, pp. 13–14, 2000.
[109] R. Price, *Droodles*, First. Simon & Schuster, 1953, p. 93.
[110] S. A. Ali, "Designing child robot interaction for facilitating creative learning," Doctoral Dissertation, Massachusetts Institute of Technology, 2019.
[111] S. Ali, N. Devasia, H. W. Park, and C. Breazeal, "Social robots as creativity eliciting agents," *Frontiers in Robotics and AI*, vol. 8, p. 673 730, 2021.
[112] S. Ali, N. Devasia, and C. Breazeal, "Designing games for enabling co-creation with social agents," *arXiv preprint arXiv:2105.00201*, 2021.
[113] P. H. Kahn Jr, B. Friedman, B. Gill, *et al.*, "A plasma display window?—the shifting baseline problem in a technologically mediated natural world," *Journal of Environmental Psychology*, vol. 28, no. 2, pp. 192–199, 2008.

[114] M. Elgarf, G. Skantze, and C. Peters, "Once upon a story: Can a creative storyteller robot stimulate creativity in children?" In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, ACM, 2021, pp. 60–67.
[115] Tallfellow Press, *Droodles compendium*, `https://tallfellow.com/droodles-compendium/`, 2025.
[116] Aalborg University. "Child protection policy at aau." (2023), [Online]. Available: `https://www.staff.aau.dk/rules/organization/child-protection-policy-pa-aau#purpose-and-scope-`.
[117] K. H. Kim, "Can we trust creativity tests? a review of the torrance tests of creative thinking (ttct)," *Creativity Research Journal*, vol. 18, no. 1, pp. 3–14, 2006.
[118] L. Rababah, "An adapted version of torrance test of creative thinking (ttct) in efl/esl writing: A rubric scoring and a review of studies," *International Journal of English and Education*, vol. 7, no. 2, p. 128, 2018.
[119] K. Richardson, *An anthropology of robots and ai: Annihilation anxiety and machines*. Routledge, Taylor & Francis Group, 2017.
[120] J. M. Jordan, *Robots*. MIT Press, 2016.
[121] K. Hornbæk and M. Hertzum, "Technology acceptance and user experience: A review of the experiential component in hci," *ACM Trans. Comput.-Hum. Interact.*, vol. 24, no. 5, Oct. 2017.
[122] E. Coronado, T. Kiyokawa, G. A. G. Ricardez, I. G. Ramirez-Alpizar, G. Venture, and N. Yamanobe, "Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0," *Journal of Manufacturing Systems*, vol. 63, pp. 392–410, 2022.
[123] Jun. 2024.
[124] G. Gasser, "The dawn of social robots: Anthropological and ethical issues," *Minds and Machines*, vol. 31, no. 3, pp. 329–336, Sep. 2021.
[125] A. Elliott, *The Routledge Social Science Handbook of Ai*. Routledge, 2022.
[126] M. H. Bruun, A. Wahlberg, R. Douglas-Jones, *et al.*, *The Palgrave Handbook of the Anthropology of Technology*. Palgrave Macmillan, 2022.
[127] M. H. Bruun, S. Hanghøj, and C. Hasse, "Studying social robots in practiced places," *Techné: Research in Philosophy and Technology*, vol. 19, no. 2, pp. 143–165, 2015.
[128] N. C. M. Nickelsen Niels Christian, "Feeding assistive robotics, socio technological imaginaries, and care: The case of bestic," *Frontiers in Artificial Intelligence and Applications*, 2018.
[129] L. Blond, "Dances with robots: Understanding social robots in practice," 2018.
[130] L. Blond and F. Olesen, "Unpacking the cultural baggage of travelling robots: How socially assistive robots are integrated in practice," in *Designing robots, designing humans*, Routledge, 2019, pp. 111–131.
[131] C. Hasse, "How robots challenge institutional practices," *Learning, Culture and Social Interaction*, vol. 26, p. 100 223, 2020.
[132] C. Breazeal, "Toward sociable robots," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 167–175, 2003.
[133] L. Fortunati, A. Esposito, M. Sarrica, and G. Ferrin, "Children's knowledge and imaginary about robots," *International Journal of Social Robotics*, vol. 7, no. 5, pp. 685–695, Aug. 2015.

[134] P. H. Kahn, T. Kanda, H. Ishiguro, *et al.*, ""robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot.," *Developmental Psychology*, vol. 48, no. 2, pp. 303–314, Mar. 2012.

[135] J. Robertson, *Robo sapiens japanicus: Robots, gender, family, and the Japanese nation*. Univ of California Press, 2018.

[136] F. KAPLAN, "Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots," *International Journal of Humanoid Robotics*, vol. 01, no. 03, pp. 465–480, 2004. eprint: `https://doi.org/10.1142/S0219843604000289`.

[137] K. Darling, *The New Breed: What our history with animals reveals about our future with robots*. Henry Hold and Company, 2021.

[138] B. Scassellati, L. Boccanfuso, C.-M. Huang, *et al.*, "Improving social skills in children with asd using a long-term, in-home social robot," *Science Robotics*, vol. 3, no. 21, Aug. 2018.

[139] UNICEF, *Children's rights and business principles*, 2012.

[140] H. Kozima, M. P. Michalowski, and C. Nakagawa, "Keepon: A playful robot for research, therapy, and entertainment," *International Journal of social robotics*, vol. 1, pp. 3–18, 2009.