ERC Starting Grant 2025
Part B1
(Part B1 is evaluated in Step 1 and Step 2, Part B2 is evaluated in Step 2 only)

Amplifying the mind with interactive AI
AmplifAI

Robin Welsch
Aalto University, Espoo, Finland
Duration: 60 months

Executive summary

Imagine a world where the limits of human intelligence are stepping stones to greater potential. By integrating artificial intelligence (AI) systems, we can amplify and expand our cognitive abilities beyond current boundaries. Yet, there is a challenge of effectively assessing and self-assessing the combined performance of humans and AI to fully harness this potential. My project will lay the groundwork for enhancing human intellect through AI integration by addressing three ambitious objectives:
(1) making AI-augmented intelligence measurable through comprehensive human-AI assessment;
(2) unconstraining metacognition in human-AI interaction to improve self-assessment and empower users;
(3) building AI interfaces with a metacognitive interaction model in an ability-centric design process.
To achieve these goals, the AmplifAI project will undertake three interlinked work packages. WP1 focuses on conceptualizing and measuring AI-augmented intelligence by developing tasks and metrics to assess human-AI composite performance, conducting large-scale testing, and exploring neural correlates via neuroimaging studies. WP2 aims to understand and remove metacognitive constraints by studying how users assess their performance with AI assistance, using Brain-Computer interfaces to track metacognitive processes, and developing adaptive interfaces to enhance self-awareness in decision-making. WP3 involves creating new AI interfaces with a metacognitive interaction model, iteratively testing prototypes, and launching the AmplifAI Challenge. Ultimately, I envision AI seamlessly integrated into human cognition, maximizing user control and enhancing capabilities. By making both assessment and self-assessment in human-AI interactions possible, we can set the stage for breakthroughs in human-AI synergy and empower people to solve complex global challenges with AI.

a Extended Synopsis

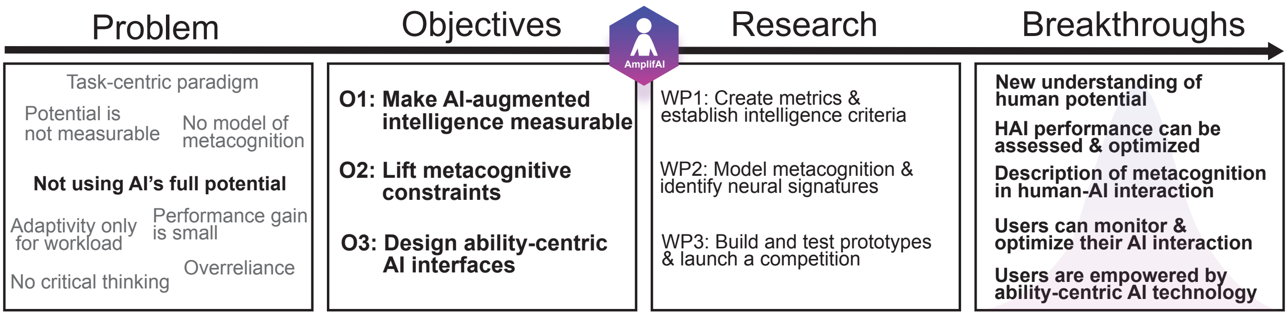

Figure 1: AmpliAI progress from the current AI problems to breakthroughs in Human-AI Interaction (HAI).

The vision of AmplifAI

If successful, AmplifAI will significantly enhance human cognitive performance through seamless AI integration. AmplifAI offers methods for designers and users to measure and enhance human-AI performance. We target to deeply integrate AI into human cognitive and metacognitive processes. For designers, AmplifAI provides **precise metrics** to optimize AI systems and interfaces for better human-AI interaction in an ability-centered design process. For users, it offers **a metacognitive interaction model to track and refine AI interactions**, enhancing task performance and effectiveness. Individuals will **transcend traditional cognitive limits**, achieving greater reasoning, decision-making, and problem-solving efficiency. This amplification will boost productivity, optimize learning, and lead to accelerated innovation and better-informed individuals. By removing metacognitive constraints and making AI-augmented intelligence measurable, people maintain **control over their cognitive processes, critically assessing AI** interaction to optimize performance. This approach ensures AI truly amplifies the mind, enhancing human capabilities without diminishing agency. In essence, AmplifAI will redefine human performance boundaries, providing means to measure and continuously surpass them, enabling individuals and societies to tackle complex challenges more effectively.

a.1 Human Cognition Challenge and Objectives

Intelligence—the capacity to learn, reason, invent, adapt, and interact socially—is at the core of human nature [21, 24]. An inherent aspect of this intelligence is our ability to use technology. Throughout history, effective tool-use has been fundamental to our development, as evidenced by the co-evolution of cognition and social structures alongside technological advancements [32, 40, 54, 47, 36, 15]. The extended mind thesis [1] posits that we extend our mind to our environment. For example, while we might struggle to multiply large numbers in our heads, a sheet of paper allows us to handle more complex calculations. Cascading this concept to the limits of computing, we now use computers to perform billions or even trillions of calculations per second, effectively extending our minds. By integrating immense computational power, tools like AI enhance our cognitive abilities, enabling us to process information on an unprecedented scale [22]. As we stand on the brink of an AI-driven technological revolution, the synergy between human cognition and machine augmentation compels us to reconsider our cognitive boundaries. Consequently, traditional psychometric intelligence tests and AI metrics [7, 39] fall short of assessing the expanded human potential realized by integrating AI technologies.

In the quest for understanding cognitive potential in the face of AI, I identified the critical weakness of current AI interfaces: They poorly support metacognition—the awareness and regulation of one's own cognitive processes [48, 31, 30, 19]. Advancements in interactive AI technologies like large language models (LLMs) offer potential enhancements to human cognitive abilities, particularly in verbal fluency and reasoning (see Noy and Zhang [37] or Bastani et al. [3]). However, I have found that [31, 30, 51, 6] while users think they improve their performance, they do not monitor their performance in interactions with AI, which is accompanied by decreased brain signals that index with error-tracking in the electroencephalogram (EEG) [51]. Similarly, there is converging evidence in the field of explainable AI where explanations of AI outputs are often ignored [50] or introduce cognitive biases [5], limiting the user's ability to optimize interactions, impoverishing human-AI performance. **Thus, I propose that metacognition, the planning, monitoring and evaluating interaction, is the key limiting factor for realizing full human-AI composite potential.**

By integrating advancements in AI with metacognitive psychology and human-computer interaction (HCI) principles, we are positioned to make groundbreaking steps in augmenting human intellect with technology. Harnessing AI's full potential in human-AI interaction requires making AI-augmented intelligence measurable and overcoming metacognitive limitations. With this integration, I hypothesize that we can create AI technologies that truly amplify the human mind.

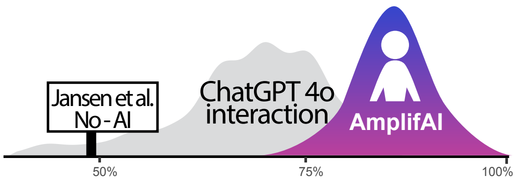

Figure 2: No-AI from Jansen et al. [26], interaction with ChatGPT-4o (our sample), and anticipated performance with design model from AmplifAI (cognitive performance ranging from 0% to 100%).

In AmplifAI, I define AI-augmented intelligence, combining Gignac [20] and Engelbart [15], as:

The maximal capacity to achieve a novel goal successfully using perceptual-cognitive interaction with computational algorithms through human-AI interfaces.

Note that, in this working definition, **maximum capacity is assumed to be a function of the interaction with the AI system**.The maximum cognitive capacity $C_{\text{max, i}}(H + A)$ of the human-AI composite for an ability domain, i, can be expressed as:

$$C_{\text{max}, i}(H + A) = \eta(\alpha, \beta) \cdot I(H; A) + \alpha \cdot H(H) + \beta \cdot H(A) - \lambda(\alpha, \beta), \quad (1)$$

where $\eta(\alpha, \beta)$ represents the efficiency of the human-AI interaction as a function of both human metacognition α and AI processing β, reflecting the AI's ability to process information and support the human, $\lambda(\alpha, \beta)$ represents the cognitive resources or load required for the human to effectively use the AI, which is a function of both α and β. Based on this working definition, we can establish AI-augmented intelligence as conceptually distinct from classical human cognitive performance, which is estimated without any technology. Breaking down this definition further, we can develop the three ambitious objectives of AmplifAI.

First, to develop AI-augmented intelligence, we need to make it, $C_{\text{max}, i}(H + A)$, measurable (**Objective: O1**). Currently, there is no conceptual basis for augmenting human intelligence with AI, which means we cannot measure or estimate the maximum capacity of the human-AI composite. **Establishing new metric**s will allow AI designers to optimize human-AI performance and understand its full potential.

Second, we need to unconstrain metacognition in human-AI interaction (**O2**). The way we think about thinking machines [48] is key to **solving challenges like lack of agency, deskilling, overreliance and trust in human-AI interaction** and, thus, harnessing AI's full potential. By researching perceptual-cognitive AI interaction processes, $\eta(\alpha, \beta)$, from a metacognitive perspective, we can enhance users' awareness and regulation of their cognitive processes when interacting with AI, ultimately empowering users to level up their cognitive abilities.

Third, many current human-centred AI designs, like in hybrid intelligence engineering, limit AI performance by developing specific applications, e.g., see [2, 25, 28, 56]. This approach overlooks the foundational characteristics of current AI systems like LLMs, restricting their broader potential. Instead of confining AI to narrow tasks, we must empower users to extend the applicability of foundational models.

To achieve this, we must **build AI interfaces for each ability i, in an ability-centric design process** (**O3**; see Figure 3) to increase our capacity to pursue novel goals and problems. Designing AI interfaces as artefacts that are optimal for improving cognitive abilities rather than fixing task-specific issues will **enable users to make use of AI systems with user control**.

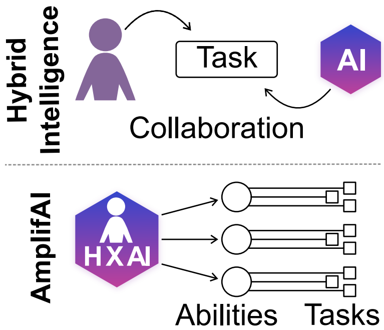

Figure 3: Hybrid intelligence is task-specific, with task-driven human-centered design. AmplifAi makes human-AI interaction (H X AI) ability-centric, empowering users.

a.2 State-of-the-Art and Beyond

Interacting with AI can redefine human cognitive potential. Non-technological attempts to enhance intelligence—such as diet, ambient color, or music—have minimal effects [4, 53, 46, 23], whereas engaging with AI technologies yields cognitive improvements. For instance, LLMs have increased essay writing speed and quality by 40% and 18%, respectively [37], and reinforcement learning-based interfaces have boosted multitasking performance to superhuman levels [34]. Thus, human-AI interaction prompts us to reconsider the fundamental goal of HCI to augment human intellect, as outlined by Engelbart and others [15, 42], though we currently lack the means to measure this composite potential.

If human-AI interaction design is successful, AI performance becomes the new baseline for human-AI potential. My preliminary data from participants taking the Law-School-Admission Test (LSAT) —a standardized test widely administered in the U.S.— with ChatGPT-4 support show that participants increased their performance by about 50% when collaborating with AI, and about half surpassed the average LSAT performance of ChatGPT-4o. However, as illustrated in Figure 4 and Figure 2, we have yet to achieve our full potential.

The explanation as to why participants do not level up more or some might even perform worse lies in Figure 4: users largely overestimate their performance and do not monitor their interactions. Lack of appropriate monitoring during AI interaction, as shown in our recent study [30] on interacting with sham-AI systems, can hinder improvement (see [38] for a computational perspective on interaction design as optimization). Moreover, this lack of metacognitive awareness not only limits performance gains but also leads to a lack of agency in interaction as showcased in our study on publishing AI-created content [14]. This aligns with findings that people using AI are overconfident and feel diminished responsibility when using AI for decision-making or content creation [12, 14, 3, 29]. While metacognitive awareness can be addressed through calibration for specific tasks, our large-scale survey found that foundational AI models like LLMs are appropriated by users for diverse tasks in different contexts [13]. Therefore, task-based human-centered design will only offer local solutions for human-AI interaction.

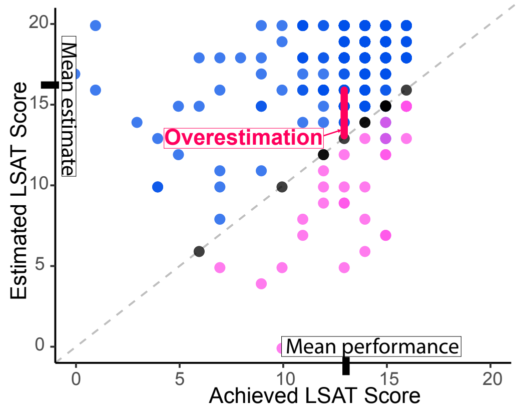

Figure 4: Scatter plot of a pilot study (N = 246) of people using ChatGPT-4 to complete the logical reasoning part of the Law School Admission Test (LSAT), showing that people do not monitor their performance - unrelated to AI-literacy.

A more general solution would be to acknowledge the foundational aspect of current AI models like LLMs and adopt an ability-centered design process [55]. This approach focuses on enhancing abilities within a domain rather than on task performance, usability, or user experience. See Figure 3 for a comparison of AmplifAI to the prominent task-based design of human-centered AI and hybrid intelligence [44]. With an ability-centered approach, I hypothesize we can more reliably solve a broad range of challenges with AI and design human-AI interactions where task-driven hybrid intelligence and human-centered AI interface design reach their limits.

By advancing beyond the current state-of-the-art with a clear and measurable concept of AI-augmented intelligence, a metacognitive interaction model, and an ability-centered design process, I hypothesize that we can build technologies that redefine the boundaries of intelligence.

Methodology

My team and I will conduct a comprehensive series of studies, including expert consultations, observational studies, laboratory experiments, and large-scale online experiments. This ensures robust quantitative data alongside deep qualitative insights into cognition and metacognition in human-AI interaction, aiming to build new AI interfaces that truly amplify our minds. We are uniquely positioned to tackle the objectives of AmplifAI as we have worked with methods of psychometrics [52], cognitive modelling [30], metacognition [14] and neuroimaging [51] for human-AI interaction. With this methodological basis, we can approach the measurement and improvement of human-AI composite performance from a new angle.

a.3 Work Packages

The project consists of three interlinked work packages (WPs) with distinct Key Intermediate goals (see Figure 5), each advancing specific goals of AmplifAI (see Figure 1). Initially, I will use LLMs as examples of foundational models in WP1 and WP2. In WP3, I will move beyond LLMs to validate our insights in human-AI interaction with other foundational models (see Figure 6), ensuring an ability-centered design approach [55].

For the project, two new PhD students, one with a focus on neuroscience/ psychology and HCI one with a focus on AI/Machine Learning and statistics, will join the Engineering Psychology Group, as well as one or two post-doctoral researchers (overall 3 years that could be split into two positions). For their contribution to each WP, see Figure 5.

a.3.1 WP1: Conceptualizing and Measuring AI-Augmented Intelligence

To establish AI-augmented intelligence as a construct and make it measurable, my team and I will use three classical criteria (conceptual, correlational, and developmental) for establishing an intelligence domain [35]. First, for the conceptual criterion, I will develop a comprehensive set of metrics covering cognitive abilities such as visual search (e.g., how many edible berries can be picked in a forest in a given time) and decision-

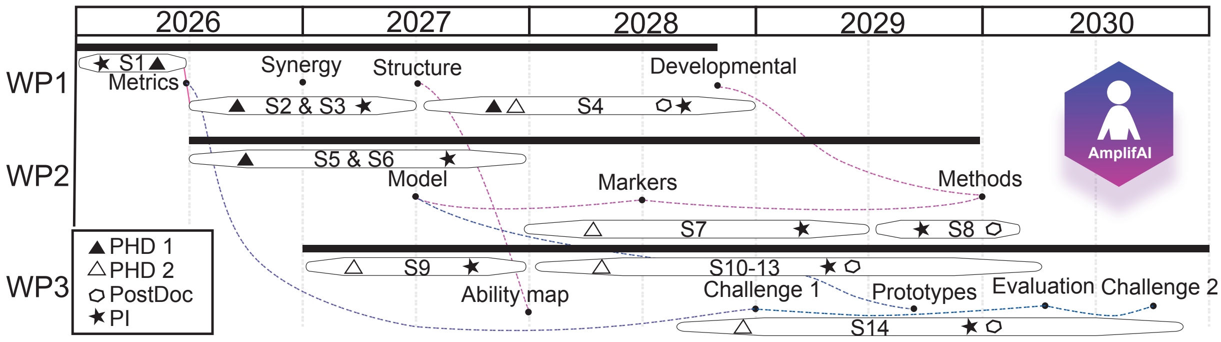

Figure 5: GANT-Chart with Work packages (WP) and Key Intermediate Goals ○. Lines connect Key Intermediate goals.

making (e.g., complex route planning), using expert interviews and a Delphi method to consolidate metrics and define them [8, 45, 18]. . I will test a large sample of participants (based on simulation/power analysis) with and without AI assistance to measure the synergetic effects of human-AI collaboration in a large-scale study, with some participants using LLMs and others not. If the group of users with the LLM improves their performance above and beyond the user group without the LLM alone across tasks, we can establish an AI-augmented intelligence as a distinct theoretical construct (**Hypothesis; H1.1: Interactive AI can improve cognitive performance.**).

Second, to test the correlational criterion, I will use factor analysis to determine if augmented intelligence with AI exhibits a distinct factorial structure compared to traditional human intelligence models like the Cattell-Horn-Carroll (CHC) theory [43, 41]. This involves analyzing whether factors associated with AI-augmented abilities correlate more strongly among themselves than with traditional cognitive abilities (**H1.2: Using AI can change the structure of intelligence.**). We will confirm the structure in another large-scale sample and test its psychometric properties. We will control for AI literacy (for the importance of AI literacy, see Article 4 of the EU AI Act) throughout our studies [33]. However, our preliminary data shows that it is unrelated to human-AI performance or metacognition [17].

Third, to address the developmental criterion, I will conduct a neuroimaging (fMRI) study to investigate how prolonged use of AI affects brain function. High-frequency and low-frequency users of AI language models will be compared during cognitive tasks to observe functional brain changes associated with AI-augmented intelligence [14, 16] (**H1.3: Prolonged use of AI leads to specific changes in brain function.**). To ensure the success of the study, we will collaborate with experts in Neuroscience from Aalto Neuroscience.

In WP1, we test AI-augmented intelligence against established criteria to confirm it as a distinct domain in intelligence. With this, we will set a new theoretical and empirical foundation for designers to improve AI interfaces measurably and renew research on intelligence in the face of AI.

a.3.2 WP2: Understanding Metacognition in AI-Augmented Intelligence

I will investigate the impact of AI interaction on metacognitive accuracy and sensitivity by having participants complete cognitive tasks with and without AI support, measuring confidence levels, self-assessed performance and sense of agency [11, 19] (**H2.1: Metacognitive accuracy, sensitivity, and sense of agency will deteriorate when interacting with AI, and agency will correlate negatively with metacognitive sensitivity.**). This builds the foundation for our model of metacognition in human-AI interaction. From this, I will identify neural markers of metacognition through EEG monitoring during decision-making, focusing on event-related potentials linked to confidence judgments and tracking of errors [51, 27] (**H2.2.: One can detect neuronal signatures of metacognitive processing during AI interaction.**). I will investigate if, with Brain-computer interfaces (BCI), one could detect moments of low metacognitive monitoring in real time, which can inform adaptive interface design. I will test strategies to enhance metacognitive monitoring, such as confidence calibration and prompting critical reflection [19, 48], and then develop adaptive interfaces that adjust based on detected metacognitive states, promoting better alignment between confidence and performance (**H2.3:Real-time interface adaptation can promote critical thinking in human-AI interaction.**). The technical challenge lies in adapting AI interaction in real-time based on metacognitive states. We will start early by implementing a Brain-Computer interface adaptation pipeline similar to Chiossi et al. [10] and then increase complexity by integrating EEG.

WP2 will produce a metacognitive model of human-AI interaction and concepts for enhancing self-assessment in neuro-adaptive interfaces. The outcome of WP2 is a new interaction model in human-AI interaction based on metacognition that emphasizes self-assessment and user control, ultimately empowering users to improve human-AI interaction themselves.

a.3.3 WP3: Developing Artifacts and a New Design Process

I will build new prototype technologies that amplify human cognitive abilities with AI, focused on supporting metacognition. My team and I will use an ability-centered design approach [55], aiming to enhance broad human cognitive abilities (e.g., LLM-based problem re-representation enabling a new level of reasoning) rather than just improving task-specific performance (**H3.1: AI interfaces designed using an ability-centric approach can amplify human cognitive abilities beyond task-specific improvements.**). Through design workshops with experts, we will create prototypes that map onto factors of intelligence. These prototypes will vary in interaction dimensions like adaptivity and autonomy and integrate means to sense the user and the environment [49, 9] which allows for grounding interaction in context and (meta-)cognitive states (**H3.2: Adaptive AI interfaces that sense user state and environment can significantly enhance broad cognitive abilities.**).

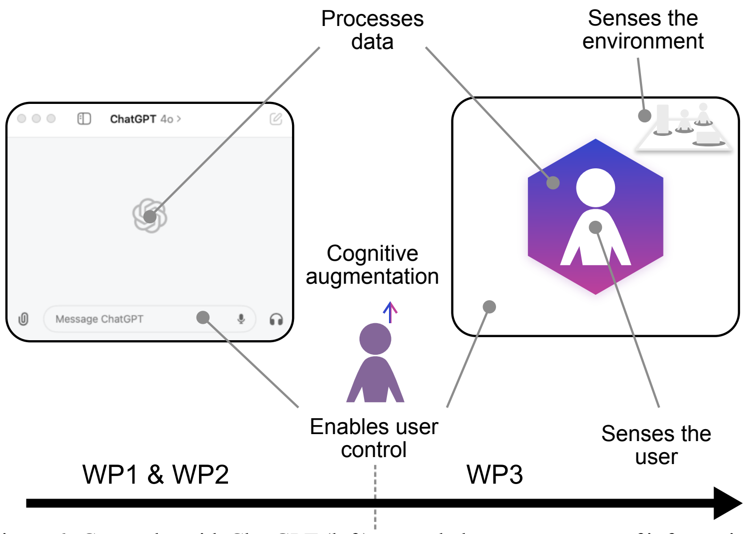

Figure 6: Currently, with ChatGPT (left), we only have a processor of information but no sensing of the user's state or environment to add context. To truly amplify the mind, we need to integrate these aspects into the interaction with foundational models like LLMs (right).

These prototypes will be tested iteratively in virtual environments, controlled lab settings, and real-world contexts. Note that with this, I will only be able to test a fixed set of AI-augmentation technologies. To scale the development and refine the design process, I will, in the third year, establish the AmplifAI Challenge, an annual competition to evaluate AI augmentation technologies in real-world scenarios. Insights from the competition will provide feedback into the broader research and development process, ensuring that the metrics, models, and strategies developed are closely aligned with real-world challenges (**H3.3: AI-augmentation technologies can be optimized in a competition of creators.**). With WP3, we have set the stage for a new design method in human-AI interaction and a new community that strives to engineer interfaces that improve human-AI composite performance. WP3 will be supported by a post-doctoral researcher with a background in prototyping and adaptive neurotechnology to guide complex software and hardware integration. Note that WP3 is a high-risk, high-reward series of technically complex studies. Anticipating potential challenges, I will hire a postdoctoral researcher specializing in prototyping neurotechnologies.

WP3 will revolutionize applied AI by establishing foundations for designing within a new interaction paradigm. It will deliver prototypes that enhance broad cognitive abilities through AI-augmented intelligence. Additionally, WP3 will initiate a new community via the AmplifAI Challenge, improving knowledge dissemination and supporting the sustainable development of AmplifAI's vision.

a.4 Impact

The AmplifAI project will significantly advance my career by establishing me as a leader in human-AI interaction and AI-based cognitive augmentation. By developing a framework that measures and enhances human-AI performance, I will lay the foundation for future research. After this project, I plan to create interactive AI systems that personalize and optimize human cognitive abilities, pushing the boundaries of AI-enhanced performance. Scientifically and technologically, AmplifAI promises to redefine human-AI synergy, leading to interfaces that actively amplify human cognition. In the next five years, this could revolutionize education, healthcare, and decision-making by enabling higher levels of efficiency, with innovations in interaction design implemented across consumer AI technologies. Looking ten years ahead, widespread AI augmentation could foster a society where enhanced cognitive abilities are commonplace, driving innovation, productivity, and improved quality of life within a vibrant research and industry environment.

References

[1] Clark Andy and Chalmers David. 1998. The extended mind. *Analysis* 58, 1 (1998), 7–19. https://doi.org/10.7551/mitpress/8535.003.0002

[2] Ian Arawjo, Chelse Swoopes, Priyan Vaithilingam, Martin Wattenberg, and Elena L. Glassman. 2024. ChainForge: A Visual Toolkit for Prompt Engineering and LLM Hypothesis Testing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, Corina Sas, Max L. Wilson, Phoebe O. Toups Dugas, and Irina Shklovski (Eds.). ACM, 304:1–304:18. https://doi.org/10.1145/3613904.3642016

[3] Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2024. Generative AI Can Harm Learning. *Available at SSRN 4895486* (2024). https://doi.org/10.2139/ssrn.4895486

[4] France Bellisle. 2004. Effects of diet on behaviour and cognition in children. *British Journal of nutrition* 92, S2 (2004), S227–S232. https://doi.org/10.1079/bjn20041171

[5] Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2022. How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review. In *AIES '22: AAAI/ACM Conference on AI, Ethics, and Society, Oxford, United Kingdom, May 19 - 21, 2021*, Vincent Conitzer, John Tasioulas, Matthias Scheutz, Ryan Calo, Martina Mara, and Annette Zimmermann (Eds.). ACM, 78–91. https://doi.org/10.1145/3514094.3534164

[6] Esther Bosch, Robin Welsch, Tamim Ayach, Christopher Katins, and Thomas Kosch. 2024. The Illusion of Performance: The Effect of Phantom Display Refresh Rates on User Expectations and Reaction Times. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, and Corina Sas (Eds.). ACM, 317:1–317:6. https://doi.org/10.1145/3613905.3650875

[7] Selmer Bringsjord. 2011. Psychometric artificial intelligence. *J. Exp. Theor. Artif. Intell.* 23, 3 (2011), 271–277. https://doi.org/10.1080/0952813X.2010.502314

[8] S Ayhan Çalışkan, Kadir Demir, and Ozan Karaca. 2022. Artificial intelligence in medical education curriculum: An e-Delphi study for competencies. *PLoS One* 17, 7 (2022), e0271872. https://doi.org/10.1371/journal.pone.0271872

[9] Francesco Chiossi, Yagiz Turgut, Robin Welsch, and Sven Mayer. 2023. Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality. *Proc. ACM Hum. Comput. Interact.* 7, MHCI (2023), 1–26. https://doi.org/10.1145/3604243

[10] Francesco Chiossi, Steeven Villa, Melanie Hauser, Robin Welsch, and Lewis L. Chuang. 2022. Design of On-body Tactile Displays to Enhance Situation Awareness in Automated Vehicles. In *IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications, CIVEMSA 2022, Chemnitz, Germany, June 15-17, 2022*. IEEE, 1–6. https://doi.org/10.1109/CIVEMSA53371.2022.9853690

[11] Clara Colombatto and Stephen Fleming. 2023. Illusions of confidence in artificial systems. (2023). https://doi.org/10.31234/osf.io/mjx2v

[12] Valdemar Danry, Pat Pataranutaporn, Florian 'Floyd' Mueller, Pattie Maes, and Sang-won Leigh. 2022. On Eliciting a Sense of Self when Integrating with Computers. In *AHs 2022: Augmented Humans 2022, Hybrid Event / Munich, Germany, March 13 - 15, 2022*, Masaaki Mochimaru, Yuichi Kurita, Martin Schmitz, and Karola Marky (Eds.). ACM, 68–81. https://doi.org/10.1145/3519391.3519414

[13] Fiona Draxler, Daniel Buschek, Mikke Tavast, Perttu Hämäläinen, Albrecht Schmidt, Juhi Kulshrestha, and Robin Welsch. 2023. Gender, Age, and Technology Education Influence the Adoption and Appropriation of LLMs. *CoRR* abs/2310.06556 (2023). https://doi.org/10.48550/ARXIV.2310.06556 arXiv:2310.06556

[14] Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. The AI Ghostwriter Effect: When Users do not Perceive Ownership of AI-Generated Text but Self-Declare as Authors. *ACM Trans. Comput. Hum. Interact.* 31, 2 (2024), 25:1–25:40. https://doi.org/10.1145/3637875

[15] Douglas C Engelbart. 1962. Augmenting human intellect: A conceptual framework. *Menlo Park, CA* (1962), 21.

[16] Iveta Fajnerová, David Greguš, Jaroslav Hlinka, Tereza Nekovářová, Antonín Škoch, Tomáš Zítka, Jan Romportl, Eva Žáčková, and Jiří Horáček. 2018. Could prolonged usage of GPS navigation implemented in augmented reality smart glasses affect hippocampal functional connectivity? *BioMed research international* 2018, 1 (2018), 2716134. https://doi.org/10.1155/2018/2716134

[17] Daniela Fernandes, Steeven Villa, Salla Nicholls, Otso Haavisto, Daniel Buschek, Albrecht Schmidt, Thomas Kosch, Chenxinran Shen, and Robin Welsch. 2024. AI Makes You Smarter, But None The Wiser: The Disconnect Between Performance and Metacognition. *arXiv preprint arXiv:2409.16708* (2024). https://doi.org/10.48550/arXiv.2409.16708

[18] R Fernández-Ballesteros, EEJ De Bruyn, A Godoy, LF Hornke, J Ter Laak, C Vizcarro, K Westhoff, H Westmeyer, and JL Zaccagnini. 2001. Guidelines for the Assessment Process (GAP): A proposal for discussion. *European Journal of Psychological Assessment* 17, 3 (2001), 187.

[19] Stephen M Fleming. 2024. Metacognition and confidence: A review and synthesis. *Annual Review of Psychology* 75 (2024), 241–268. https://doi.org/10.1146/annurev-psych-022423-032425

[20] Gilles E Gignac. 2024. Rethinking the Dunning-Kruger effect: Negligible influence on a limited segment of the population. *Intelligence* 104 (2024), 101830.

[21] Gilles E Gignac and Eva T Szodorai. 2024. Defining intelligence: Bridging the gap between human and artificial perspectives. *Intelligence* 104 (2024), 101832. https://doi.org/10.1016/j.intell.2024.101832

[22] José Hernández-Orallo and Karina Vold. 2019. AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, Honolulu, HI, USA, January 27-28, 2019*, Vincent Conitzer, Gillian K. Hadfield, and Shannon Vallor (Eds.). ACM, 507–513. https://doi.org/10.1145/3306618.3314238

[23] Aron T Hill, Paul B Fitzgerald, and Kate E Hoy. 2016. Effects of anodal transcranial direct current stimulation on working memory: a systematic review and meta-analysis of findings from healthy and neuropsychiatric populations. *Brain stimulation* 9, 2 (2016), 197–208.

[24] N. K. Humphrey. 1988. The social function of intellect. In *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, R. W. Byrne and A. Whiten (Eds.). Clarendon Press/Oxford University Press, 13–26. (Reprinted from "Growing Points in Ethology", 1976, ed. P. P. G. Bateson and R. A. Hinde, pp. 303–317, Cambridge University Press).

[25] Andrés Isaza-Giraldo, Paulo Bala, Pedro F. Campos, and Lucas Pereira. 2024. Prompt-Gaming: A Pilot Study on LLM-Evaluating Agent in a Meaningful Energy Game. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, and Corina Sas (Eds.). ACM, 272:1–272:12. https://doi.org/10.1145/3613905.3650774

[26] Rachel A Jansen, Anna N Rafferty, and Thomas L Griffiths. 2021. A rational model of the Dunning–Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behaviour* 5, 6 (2021), 756–763. https://doi.org/10.1038/s41562-021-01057-0

[27] Suk Yun Kang, Chang-Hwan Im, Miseon Shim, Fatta B Nahab, Jihye Park, Do-Won Kim, John Kakareka, Nathanial Miletta, and Mark Hallett. 2015. Brain Networks Responsible for Sense of Agency: An EEG Study. *PLoS ONE* 10, 8 (2015), e0135261. https://doi.org/10.1371/journal.pone.0135261

[28] Minsol Kim, Aliea L. Nallbani, and Abby Rayne Stovall. 2024. Exploring LLM-based Chatbot for Language Learning and Cultivation of Growth Mindset. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, and Corina Sas (Eds.). ACM, 609:1–609:5. https://doi.org/10.1145/3613905.3648628

[29] Sunnie SY Kim, Q Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. " I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 822–835.

[30] Agnes Mercedes Kloft, Robin Welsch, Thomas Kosch, and Steeven Villa. 2024. "AI enhances our performance, I have no doubt this one will do the same": The Placebo effect is robust to negative descriptions of AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, Corina Sas, Max L. Wilson, Phoebe O. Toups Dugas, and Irina Shklovski (Eds.). ACM, 299:1–299:24. https://doi.org/10.1145/3613904.3642633

[31] Thomas Kosch, Robin Welsch, Lewis L. Chuang, and Albrecht Schmidt. 2022. The Placebo Effect of Artificial Intelligence in Human-Computer Interaction. *ACM Trans. Comput. Hum. Interact.* 29, 6 (2022), 56:1–56:32. https://doi.org/10.1145/3529225

[32] Kevin Laland and Amanda Seed. 2021. Understanding human cognitive uniqueness. *Annual Review of Psychology* 72, 1 (2021), 689–716. https://doi.org/10.1146/annurev-psych-062220-051256

[33] Matthias Carl Laupichler, Alexandra Aster, Nicolas Haverkamp, and Tobias Raupach. 2023. Development of the "scale for the assessment of non-experts' AI literacy"–An exploratory factor analysis. *Computers in Human Behavior Reports* 12 (2023), 100338. https://doi.org/10.1016/j.chbr.2023.100338

[34] Alexander Lingler, Dinara Talypova, Jussi P. P. Jokinen, Antti Oulasvirta, and Philipp Wintersberger. 2024. Supporting Task Switching with Reinforcement Learning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, Corina Sas, Max L. Wilson, Phoebe O. Toups Dugas, and Irina Shklovski (Eds.). ACM, 82:1–82:18. https://doi.org/10.1145/3613904.3642063

[35] John D Mayer, David R Caruso, and Peter Salovey. 1999. Emotional intelligence meets traditional standards for an intelligence. *Intelligence* 27, 4 (1999), 267–298. https://doi.org/10.1016/S0160-2896(99)00016-1

[36] Ana F Navarrete, Simon M Reader, Sally E Street, Andrew Whalen, and Kevin N Laland. 2016. The coevolution of innovation and technical intelligence in primates. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371, 1690 (2016), 20150186. https://doi.org/10.1098/rstb.2015.0186

[37] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654 (2023), 187–192. https://doi.org/10.1126/science.adh2586 arXiv:https://www.science.org/doi/pdf/10.1126/science.adh2586

[38] Antti Oulasvirta, Jussi P. P. Jokinen, and Andrew Howes. 2022. Computational Rationality as a Theory of Interaction. In *CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 - 5 May 2022*, Simone D. J. Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Mark Drucker, Julie R. Williamson, and Koji Yatani (Eds.). ACM, 359:1–359:14. https://doi.org/10.1145/3491102.3517739

[39] Max Pellert, Clemens M Lechner, Claudia Wagner, Beatrice Rammstedt, and Markus Strohmaier. 2023. Ai psychometrics: Assessing the psychological profiles of large language models through psychometric inventories. *Perspectives on Psychological Science* (2023), 17456916231214460. https://doi.org/10.1177/17456916231214460

[40] Steven Pinker. 2010. The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences* 107 (2010), 8993–8999. https://doi.org/10.1073/pnas.0914630107

[41] Alexander O Savi, Maarten Marsman, Han LJ van der Maas, and Gunter KJ Maris. 2019. The wiring of intelligence. *Perspectives on Psychological Science* 14, 6 (2019), 1034–1061. https://doi.org/10.1177/1745691619866447

[42] Albrecht Schmidt. 2017. Augmenting Human Intellect and Amplifying Perception and Cognition. *IEEE Pervasive Comput.* 16, 1 (2017), 6–10. https://doi.org/10.1109/MPRV.2017.8

[43] W Joel Schneider and Kevin S McGrew. 2018. The Cattell-Horn-Carroll theory of cognitive abilities. *Contemporary intellectual assessment: Theories, tests, and issues* (2018), 73–163.

[44] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *Int. J. Hum. Comput. Interact.* 36, 6 (2020), 495–504. https://doi.org/10.1080/10447318.2020.1741118

[45] Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, and Abu Awal Md Shoeb et al. 2023. Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. *Trans. Mach. Learn. Res.* 2023 (2023). https://openreview.net/forum?id=uyTL5Bvosj

[46] Kenneth M Steele, Simone Dalla Bella, Isabelle Peretz, Tracey Dunlop, Lloyd A Dawe, G Keith Humphrey, Roberta A Shannon, Johnny L Kirby Jr, and CG Olmstead. 1999. Prelude or requiem for the 'Mozart effect'? *Nature* 400, 6747 (1999), 827–827. https://doi.org/10.1016/j.brs.2015.10.006

[47] Sally E Street, Ana F Navarrete, Simon M Reader, and Kevin N Laland. 2017. Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proceedings of the National Academy of Sciences* 114, 30 (2017), 7908–7914. https://doi.org/10.1073/pnas.1620734114

[48] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, Corina Sas, Max L. Wilson, Phoebe O. Toups Dugas, and Irina Shklovski (Eds.). ACM, 680:1–680:24. https://doi.org/10.1145/3613904.3642902

[49] Niels van Berkel, Mikael B. Skov, and Jesper Kjeldskov. 2021. Human-AI interaction: intermittent, continuous, and proactive. *Interactions* 28, 6 (2021), 67–71. https://doi.org/10.1145/3486941

[50] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S. Bernstein, and Ranjay Krishna. 2023. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *Proc. ACM Hum. Comput. Interact.* 7, CSCW1 (2023), 1–38. https://doi.org/10.1145/3579605

[51] Steeven Villa, Thomas Kosch, Felix Grelka, Albrecht Schmidt, and Robin Welsch. 2023. The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Comput. Hum. Behav.* 146 (2023), 107787. https://doi.org/10.1016/J.CHB.2023.107787

[52] Steeven Villa, Jasmin Niess, Albrecht Schmidt, and Robin Welsch. 2023. Society's Attitudes Towards Human Augmentation and Performance Enhancement Technologies (SHAPE) Scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3 (2023), 128:1–128:23. https://doi.org/10.1145/3610915

[53] Christoph von Castell, Daniela Stelzmann, Daniel Oberfeld, Robin Welsch, and Heiko Hecht. 2018. Cognitive performance and emotion are indifferent to ambient color. *Color Research & Application* 43, 1 (2018), 65–74. https://doi.org/10.1002/col.22168

[54] Andrew Whiten and David Erdal. 2012. The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, 1599 (2012), 2119–2129. https://doi.org/10.1098/rstb.2012.0114

[55] Jacob O. Wobbrock, Shaun K. Kane, Krzysztof Z. Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-Based Design: Concept, Principles and Examples. *ACM Trans. Access. Comput.* 3, 3 (2011), 9:1–9:27. https://doi.org/10.1145/1952383.1952384

[56] Ruican Zhong, Donghoon Shin, Rosemary Meza, Predrag Klasnja, Lucas Colusso, and Gary Hsieh. 2024. AI-Assisted Causal Pathway Diagram for Human-Centered Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11-16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, Corina Sas, Max L. Wilson, Phoebe O. Toups Dugas, and Irina Shklovski (Eds.). ACM, 2:1–2:19. https://doi.org/10.1145/3613904.3642179