

A Inteligência Artificial (IA) vai perceber: Explorar a relação entre cognição híbrida, confiança e incerteza em Interfaces Abertas de IA Generativa

Paulo Ferreira

Tallinn University

paulojqf@tlu.ee

Sónia Sousa

Faculdade de Engenharia da Universidade do Porto

scsousa@fe.up.pt

Resumo: As interfaces dos sistemas comerciais de IA generativa de uso generalizado, como o *ChatGPT*, *Claude* ou *Gemini*, podem influenciar a forma como o esforço cognitivo é distribuído antes e durante a execução de tarefas. Este artigo propõe dois conceitos para descrever essa dinâmica: compensação de sistema, entendida como a capacidade do sistema para gerar respostas coerentes a partir de inputs subespecificados, e delegação da intenção, entendida como o comportamento do utilizador de fornecer *prompts* incompletas e confiar no sistema para inferir parâmetros em falta. Argumenta-se que a compensação de sistema pode enfraquecer sinais de calibração que, noutros contextos de interação, incentivariam uma formulação mais deliberada da intenção. Com base em quatro sessões exploratórias com 33 participantes, identificaram-se cinco padrões recorrentes: (1) subespecificação estruturada, (2) simplificação após o output do sistema, (3) passagem do refinamento do input para o ajustamento do output, (4) reenquadramento progressivo da tarefa e (5) discrepância entre produção e avaliação de *prompts*. Os resultados sugerem que, em contextos de aprendizagem, os utilizadores tendem a delegar no sistema parte da especificação inicial da tarefa, embora mantenham capacidade para reconhecer *prompts* mais bem formulados quando lhes são apresentados. Discute-se ainda o papel da fricção construtiva como possível estratégia de desenho de interface para apoiar uma interação humano-IA mais deliberada e calibrada.

Palavras-chave: Confiança na tecnologia, Interação humano-Inteligência Artificial, Design de Interação, Interação Humano-Computador, Experiência do Utilizador.

Abstract: The interfaces of widely used, general-purpose generative AI systems such as ChatGPT, Claude, or Gemini can influence how cognitive effort is distributed before and during task execution. This article proposes two concepts to describe this dynamic: system compensation, understood as the system's ability to generate coherent responses from underspecified inputs, and intention delegation, understood as the user's tendency to provide incomplete prompts and rely on the system to infer missing parameters. The article argues that system compensation can weaken calibration cues that, in other interaction contexts, would encourage more deliberate intention formulation. Based on four exploratory sessions with 33 participants, five recurring patterns were identified: (1) structured underspecification, (2) simplification after the system's output, (3) a shift from input refinement to output adjustment, (4) progressive task reframing, and (5) a mismatch between prompt production and prompt evaluation. The results suggest that, in learning contexts, users tend to delegate part of the initial task specification to the system, although they retain the ability to recognize better-formulated prompts when these are presented to them. The article also discusses the role of constructive friction as a possible interface design strategy to support more deliberate and calibrated human-AI interaction.

Keywords: Trustworthy AI, Human-AI interaction, Interaction Design, Human Computer Interaction, User experience.

1. Introdução

Os sistemas comerciais de IA generativa (GenAI) introduzem um paradigma de interação distinto das interfaces gráficas tradicionais: em vez de opções e controlos predefinidos, oferecem um campo aberto de texto/voz que incentiva pedidos em linguagem natural ("pergunta-me qualquer coisa"). Embora esta abertura aumente a flexibilidade e a facilidade de uso, também pode estar a mudar a forma como os utilizadores formulam tarefas, porque exige que expressem a sua intenção antes da ação, enquanto o sistema consegue gerar respostas coerentes mesmo a partir de *prompts* subespecificadas (Dalsgaard, 2025). A partir da ergonomia cognitiva (Hollnagel, 1997), o artigo define esta capacidade como compensação de sistema: o sistema infere restrições e parâmetros que o utilizador não explicita (Vaswani et al., 2017).

Com base nisso, investiga como interfaces abertas de GenAI influenciam a distribuição do trabalho cognitivo em contextos de aprendizagem, introduzindo o conceito de delegação da intenção, que descreve quando o utilizador “entrega” ao sistema parte da especificação da tarefa através de *prompts* incompletas.

O estudo segue uma abordagem de Research-through-Design (RtD) (Zimmerman et al., 2007) e inclui quatro sessões exploratórias com investigadores, docentes e estudantes, analisando (1) como as *prompts* iniciais são construídas para aprender um conceito novo e complexo e (2) como evoluem ao longo de rondas sucessivas de interação. A partir das observações, o artigo discute padrões de interação associados à compensação do sistema e ao design de interfaces abertas. Por fim, propõe a fricção construtiva, inspirada nas “dificuldades desejáveis” (Bjork et al., 2011), como estratégia de design para incentivar uma formulação mais deliberada da intenção antes da resposta, promovendo uma interação humano-IA mais reflexiva e cognitivamente negociada.

2. O problema e a sua contextualização

O problema investigado é que as interfaces abertas de GenAI, ao gerarem respostas plausíveis mesmo a partir de *prompts* incompletas, podem reduzir a necessidade percebida de formular a intenção inicial com precisão. Em contextos educativos, isso pode levar os utilizadores a delegar parte da especificação cognitiva da tarefa para o sistema, com possíveis implicações para a metacognição, a aprendizagem e a agência humana. Assim, este artigo contribui com: (1) uma conceptualização inicial de compensação do sistema e delegação da intenção no uso de GenAI; (2) observações empíricas sobre a evolução das *prompts* em diferentes configurações de interação; e (3) sugestões exploratórias de design de interface para promover uma interação humano-IA mais equilibrada em sistemas generativos abertos de uso generalizado.

1.1. A *Prompt* como Forma de Interação no Golfo de Execução

Do ponto de vista do utilizador, a interface é o meio através do qual a interação acontece (Norman, 1984, 1986). Na Interação Humano-Computador (IHC), essa interação pode ser descrita por dois desafios: o Golfo de Execução, isto é, a distância entre o objetivo do utilizador e as ações necessárias para o alcançar no sistema; e o Golfo de Avaliação, a distância entre o output do sistema e a capacidade do utilizador para o interpretar. Em sistemas tradicionais, menus, botões e opções predefinidas ajudam a orientar a escolha e a especificação das ações. Já nos sistemas de IA generativa, a interação é apresentada como aberta, através de um

campo em branco onde o utilizador deve articular a sua intenção em linguagem natural por meio de uma *prompt* (Dalsgaard, 2025). Esta mudança desloca para o utilizador o trabalho inicial de especificar o que quer - agora sem o mesmo “andamento” imposto por uma interface estruturada - e pode alterar a forma como o esforço cognitivo é distribuído entre humano e sistema. À luz da cognição distribuída (Hutchins, 1995; Hollan et al., 2000), parte desse trabalho pode ser redistribuída para o sistema através do que designamos por compensação de sistema. Por isso, investigações recentes têm explorado interfaces que estruturam melhor a relação entre intenção, ação e resposta, apoiando a explicitação e o refinamento das *prompts* (Ding, 2024; Gmeiner et al., 2025; Riche et al., 2025).

Uma característica definidora dos sistemas atuais de IA generativa é a sua capacidade de produzir outputs confiantes e aparentemente competentes mesmo quando o input do utilizador é subespecificado. Em vez de exigir instruções precisas, estes sistemas inferem elementos em falta através de geração probabilística (Vaswani et al., 2017). Neste artigo, compensação de sistema refere-se à capacidade do sistema para gerar respostas preenchendo restrições que o utilizador não explicitou - por exemplo, o âmbito, o nível de detalhe, o formato ou o público-alvo. Em paradigmas tradicionais, inputs incompletos tenderiam a gerar erros ou pedidos de correção, alargando o Golfo de Execução. Em contraste, nos sistemas generativos, normalmente não há erro devolvido ao utilizador (um forte sinal de “eficiência”), porque o sistema compensa a falta de especificação com uma resposta aceitável. Isso reduz as consequências percebidas de um *input* incompleto e pode atenuar sinais de *feedback* que, noutros sistemas, ajudariam o utilizador a calibrar a necessidade de especificar melhor o pedido. Assim, a delegação da intenção pode não refletir apenas falta de literacia em IA (Annapureddy et al., 2025), mas antes uma adaptação estratégica a um ambiente de interação que a torna possível.

Quando as restrições da interface deixam de orientar a interação, como nas interfaces abertas de IA generativa, as fases de formação e seleção da intenção podem ser reconfiguradas. Segundo Norman (1986), a intenção é a representação mental do objetivo e a seleção é a tradução dessa intenção numa ação possível; este processo é dinâmico, pode avançar e recuar e pode gerar subintenções ao longo da interação. Em sistemas tradicionais, o utilizador revê operações disponíveis e especifica uma sequência de ações. Em GenAI, porém, é possível obter *outputs* longos e “completos” a partir de *prompts* pouco detalhadas, o que pode levar os utilizadores a seguir o caminho de menor resistência (Risko & Gilbert, 2016). Com base no conceito de compensação de sistema, propõe-se que os utilizadores passem a delegar a intenção: em vez de explicitar aspetos essenciais da

tarefa no *input*, deixam que o sistema infira e complete parâmetros em falta. Essa delegação pode ser estratégica (deliberada) ou estrutural (emergente do próprio ambiente de interação). O resultado é um possível desalinhamento na superação do Golfo de Execução: o utilizador inicia a interação, mas parte da especificação passa para o sistema, aumentando a distância entre a especificidade do *input* e a qualidade do *output* e, potencialmente, também o esforço cognitivo necessário para avaliar e lidar com respostas baseadas em inferências implícitas.

1.2. Calibração Cognitiva

Se os utilizadores obtêm *outputs* fluentes mesmo quando a *prompt* inicial está incompleta, isso pode reforçar a expectativa de que não é necessário fornecer um *input* detalhado. Com o tempo, essa experiência pode enfraquecer a capacidade de reconhecer quando é preciso especificar melhor o pedido (Fernandes et al., 2025). Tankelevitch et al. (2024) argumentam que a interação com sistemas de IA generativa impõe exigências metacognitivas elevadas, e Forshaw (2026) propõe que estes sistemas criam um paradoxo metacognitivo: podem corroer precisamente os processos de automonitorização cognitiva no momento em que eles são mais necessários. Neste enquadramento, usamos o conceito de calibração cognitiva para descrever a capacidade dos utilizadores avaliar se a sua *prompt* é adequada à tarefa. Sustentamos que a compensação de sistema pode degradar esse processo metacognitivo, levando os utilizadores a tratar *inputs* subespecificados como “bons o suficiente”.

A investigação clássica sobre enviesamento da automação (Endsley & Kiris, 1995; Parasuraman & Manzey, 2010) tem-se centrado sobretudo na fase de avaliação do *output*: quando os utilizadores confiam demasiado nos resultados automatizados, reduzem a verificação e delegam o julgamento no sistema. Neste artigo, não redefinimos o enviesamento da automação; propomos, em vez disso, alargar a sua relevância ao início da interação. Argumentamos que dinâmicas análogas de confiança excessiva podem emergir antes, durante a própria formação da intenção e a construção do *input*. Em sistemas de IA generativa, a compensação do sistema pode influenciar a forma como os utilizadores escrevem a *prompt* desde o primeiro momento, incentivando uma delegação precoce de aspetos essenciais da tarefa. Assim, a dependência excessiva (ou excesso de confiança) pode ser ativada não apenas na avaliação do que o sistema produz, mas já na forma como o utilizador define o que pede (Buçinca et al., 2021; Qiao et al., 2025). Esta extensão conceptual fornece uma base para analisar como a compensação de sistema pode moldar o comportamento do utilizador em interações abertas, desde a fase inicial de intenção e seleção.

3. Abordagem Metodológica

As interfaces abertas de IA generativa criam uma nova condição de interação: mesmo quando os utilizadores fornecem *inputs* incompletos, vagos ou pouco especificados, o sistema tende a gerar respostas coerentes, completas e aparentemente adequadas. Chamamos a este fenómeno compensação do sistema. Em contextos de aprendizagem, isto levanta um problema central: se o sistema “preenche” o que o utilizador não explicita, pode diminuir a necessidade percebida de formular a intenção de forma deliberada antes de interagir com a IA. Daí resulta a questão de investigação: de que modo a compensação do sistema molda a distribuição do trabalho cognitivo entre humano e IA durante a aprendizagem? Para responder, analisamos dois momentos da interação: (1) como os utilizadores constroem as *prompts* iniciais quando procuram aprender um conceito novo; e (2) como essas *prompts* são reformuladas ao longo de rondas sucessivas, à medida que os utilizadores reagem aos *outputs* produzidos pelo sistema.

O estudo segue uma metodologia de *Research-through-Design* (RtD) (Zimmerman et al., 2007). Na lógica RtD, a investigação evolui por iterações: parte-se de um enquadramento inicial do problema, recolhem-se perceções exploratórias e essas evidências informam a etapa seguinte, até que o problema de investigação possa ser formulado de forma mais clara e fundamentada. Este trabalho resulta de quatro iterações com diferentes participantes e tem como objetivo explorar como os utilizadores constroem a intenção inicial e ajustam *prompts* subsequentes (subintenções) ao interagir com sistemas de GenAI de interface aberta para aprender um conceito. O foco principal é identificar padrões de interação, e não medir desempenho nem estabelecer relações causais.

3.1. Desenho do Estudo

O desenho do estudo consistiu em quatro sessões exploratórias envolvendo 33 participantes de quatro grupos: investigadores de doutoramento, docentes do ensino superior, estudantes de licenciatura em Ciências da Comunicação e estudantes de licenciatura em Sociologia. Em todas as sessões foi proposta aos participantes uma tarefa de aprendizagem para a qual deviam usar um sistema de IA generativa para aprender sobre um conceito com o intuito de conseguirem explicá-lo a outras pessoas. A mesma tarefa foi utilizada para todas as sessões. Os participantes foram instruídos a utilizar um sistema de IA generativa para aprender sobre conceitos fundamentais da computação quântica e produzir uma explicação abrangente do seu funcionamento.

Os participantes foram recrutados em contextos acadêmicos: 8 investigadores de doutoramento para a sessão 1, 9 docentes do ensino superior para a sessão 2, 6 estudantes do ensino superior da licenciatura em Ciências da Comunicação para a sessão 3 e 10 estudantes do ensino superior da licenciatura em Sociologia para a sessão 4. Esta variação nos perfis dos participantes não foi controlada experimentalmente, mas permitiu a observação de padrões de interação em diferentes níveis de familiaridade com contextos de aprendizagem académica.

3.2. Procedimento

Antes de cada sessão, todos os participantes foram informados sobre o desenho do estudo, a recolha e o uso dos dados. Dada a natureza de baixo risco deste estudo exploratório e o objetivo de garantir anonimato total, não foram recolhidos consentimentos por escrito; foi obtido consentimento verbal explícito. As sessões não foram gravadas (áudio ou vídeo) e não foram recolhidas Informações Pessoalmente Identificáveis (PII) nem dados demográficos. A recolha de dados consistiu em *prompts* anonimizadas e notas de campo do investigador. Nas Sessões 2, 3 e 4, as *prompts* foram recolhidas através do Mentimeter; na sessão 1, as *prompts* foram produzidas verbalmente e registadas pelo investigador.

Sessão 1 (workshop presencial, Escola Doutoral da Estónia, fevereiro de 2026). Participaram 8 doutorandos. O procedimento seguiu quatro passos: (1) apresentação da tarefa; (2) cada participante propôs uma *prompt* verbalmente; (3) as *prompts* foram discutidas em grupo e uma foi selecionada; (4) o investigador introduziu a *prompt* selecionada no ChatGPT. O output foi projetado e lido coletivamente. Sempre que o grupo considerou o output insuficiente, iniciava-se uma nova ronda de *prompting*. Foram realizadas 3 rondas.

Sessão 2 (online, MS Teams). Participaram 9 docentes do ensino superior. A tarefa manteve-se, mas a dinâmica foi ajustada para reduzir a mediação do investigador e recolher contributos individuais. Após lerem a tarefa, os participantes submeteram *prompts* individualmente via Mentimeter. Em seguida, todas as *prompts* foram exibidas em simultâneo e cada participante votou para escolher a *prompt* a submeter ao sistema. Os resultados só foram mostrados após todos votarem. A *prompt* mais votada foi copiada para o ChatGPT e o output foi apresentado ao grupo. O processo foi repetido por 2 rondas, até existir consenso de que o output era suficiente.

Sessão 3 (presencial, estudantes do ensino superior). Participaram 6 estudantes e a sessão seguiu a mesma estrutura baseada em Mentimeter da sessão 2. Houve 3 rondas de submissão de *prompts* para a tarefa principal e 1 ronda adicional para uma segunda tarefa. Após concluírem a

primeira tarefa, foi introduzida uma segunda (edição genética em vez de computação quântica), usando um ChatGPT personalizado com uma condição de fricção construtiva: antes de gerar o output, o sistema exigia que os participantes acrescentassem detalhes (ex., público-alvo, formato e restrições). Os participantes interagiram com esta condição numa única ronda; as reações foram observadas e registadas.

Sessão 4 (presencial, licenciatura em Sociologia). Participaram 10 estudantes. A sessão repetiu a estrutura baseada em Mentimeter: submissão individual, exibição simultânea, votação do grupo e submissão da *prompt* ao sistema pelo investigador. Foram realizadas 3 rondas sobre a mesma tarefa (computação quântica). Não houve segunda tarefa nem condição de fricção construtiva. O maior tamanho do grupo (n=10) gerou um número mais elevado de *prompts* por ronda, resultando num conjunto de dados mais rico para analisar a discrepância entre produção e avaliação de *prompts*.

3.3. Recolha de Dados

A recolha de dados focou-se nas *prompts* textuais e em comportamentos de interação observáveis: (1) *prompts* individuais produzidas em cada ronda, (2) evolução dessas *prompts* ao longo das iterações, (3) *prompts* selecionadas pelo grupo (por discussão ou votação) e (4) notas de observação sobre reações e debates dos participantes. O estudo não mediu desempenho, carga cognitiva ou outras métricas quantitativas; em vez disso, a análise procurou identificar padrões recorrentes na forma como as *prompts* foram construídas, ajustadas e escolhidas, sustentando assim a formulação dos conceitos de compensação de sistema e delegação da intenção.

Como as sessões foram realizadas em grupo, a influência coletiva na convergência e seleção das *prompts* foi tratada como parte do contexto de interação. Nas Sessões 2, 3 e 4, o uso do Mentimeter permitiu analisar dois níveis: produção individual (as *prompts* submetidas antes de ver as dos outros, captando formulação espontânea) e julgamento coletivo (a *prompt* mais votada, refletindo o que o grupo considera “melhor”).

A comparação entre estes níveis fundamenta a análise da discrepância produção/avaliação. Para apoiar esta leitura, cada *prompt* foi classificada com uma rubrica de cinco dimensões de especificação: Público-alvo (A), Formato (F), Profundidade/Âmbito (D), Restrições (C) e Propósito/Objetivo (P), atribuindo 1 ponto por dimensão presente (pontuação total 0–5). Estas pontuações, atribuídas pelo investigador, funcionam como indicadores descritivos para caracterizar a evolução da especificação e quantificar a diferença entre as *prompts* individuais e a *prompt* escolhida pelo grupo.

4. Resultados

Os resultados apresentados nesta secção baseiam-se em observações de quatro sessões com diferenças relevantes no formato, tamanho do grupo e nível de mediação do investigador. Na sessão 1, as *prompts* foram construídas verbalmente e negociadas em grupo, sendo o investigador o operador da interface. Nas Sessões 2, 3 e 4, os participantes submeteram *prompts* individualmente via Mentimeter e, em seguida, votaram na *prompt* a introduzir no sistema. Estas variações estruturais limitam comparações diretas entre sessões. Assim, os padrões descritos abaixo devem ser interpretados como observações consistentes dentro de cada sessão, e não como resultados produzidos por um desenho experimental uniforme. Quando surgem semelhanças entre sessões, estas são apresentadas de forma indicativa, não conclusiva.

Ao longo das quatro sessões, observou-se frequentemente que os participantes formulavam *prompts* que combinavam objetivos gerais com pouca especificação de restrições. Era comum incluir enquadramentos de papel (por exemplo, “age como um professor”) ou intenções amplas (por exemplo, “explica de forma clara”), mas raramente se definia explicitamente o âmbito, a profundidade, a estrutura esperada ou critérios de avaliação do output.

Designamos este padrão como subespecificação estruturada: as *prompts* parecem organizadas em torno de alguns parâmetros visíveis (papel, tom, intenção), mas deixam aspetos essenciais da intenção por definir, delegando no sistema a tarefa de os preencher. Por exemplo, na sessão 1, um participante propôs: “Explica-me o que é a computação quântica em termos simples e adiciona referências. Assume que não sei nada.” Embora esta formulação introduza um limite relativo ao conhecimento prévio (“assume que não sei nada”), não especifica como a explicação deve ser estruturada, qual o nível de detalhe apropriado, nem que tipo de referências seriam relevantes. De forma semelhante, surgiram *prompts* centradas na autoridade, como “Age como o maior especialista do mundo em computação quântica.” Aqui, o utilizador define a perspetiva, mas não clarifica para quem é a explicação, como deve ser adaptada, ou qual o objetivo pedagógico concreto.

Em conjunto, estes exemplos sugerem que, quando confrontados com a tarefa de aprender um conceito novo, os participantes tenderam a confiar na capacidade do sistema para decidir elementos centrais da resposta, em vez de os explicitar durante a fase inicial de seleção da ação. Importa, no entanto, reconhecer que, na sessão 1, as *prompts* foram geradas através de

discussão e negociação coletiva, o que pode ter reduzido a variação individual e favorecido formulações mais “aceitáveis” socialmente. Assim, o padrão identificado pode refletir, em parte, um efeito de consenso de grupo e não apenas uma estratégia individual de *prompting*.

4.1. Estratégias de simplificação após exposição aos outputs do sistema

Após a exposição aos outputs do sistema, as *prompts* produzidas nas sessões 2, 3 e 4 sugerem uma tendência consistente para a simplificação. Esta tendência manifesta-se através de três padrões recorrentes.

Em primeiro lugar, surgem pedidos genéricos de simplificação, sem estrutura adicional, como “Explica em termos simples o que é a computação quântica”. Nestes casos, o objetivo inicial da tarefa - aprender suficientemente bem o conceito para o conseguir explicar a terceiros - desaparece da *prompt* e é substituído por uma reformulação genérica da insatisfação face ao output anterior.

Em segundo lugar, observam-se *prompts* que introduzem apenas uma restrição isolada, normalmente relacionada com o formato ou a extensão da resposta, como “Escreve uma explicação de um parágrafo sobre computação quântica” ou “Explica os conceitos básicos da computação quântica em 150 palavras para alguém sem conhecimento prévio sobre o tema”. Embora estas *prompts* acrescentem uma dimensão de especificação relativamente à ronda inicial, deixam outras dimensões importantes - como o propósito, a profundidade ou a estrutura explicativa - implícitas ou indefinidas.

Em terceiro lugar, identificam-se pedidos de simplificação combinados com uma definição mínima do público-alvo, como “Explica os conceitos básicos da computação quântica de forma simples e concisa, para que eu possa explicar o tema a uma criança de 10 anos”. Aqui, o público é explicitado, mas a *prompt* continua a não definir com clareza o que deve ser incluído, como deve ser organizado o conteúdo, ou o que contaria como uma explicação bem-sucedida.

Em conjunto, estes padrões mostram que, após receberem um *output* considerado demasiado complexo, os participantes tendem a reagir sobre as características visíveis da resposta - simplicidade, extensão, tom ou adequação ao público - em lugar de regressarem ao objetivo original da tarefa para especificarem novamente a sua intenção. Esta observação é consistente com a hipótese de delegação da intenção: em lugar de reformularem a tarefa a partir do seu propósito inicial, os participantes passam a tratar a simplificação do output como um substituto funcional do progresso da aprendizagem.

A sessão 4 reforça este padrão de forma particularmente clara. Com dez participantes a submeter *prompts* por ronda, a passagem da ronda 1 para a ronda 2 torna-se simultaneamente mais visível e mais diversificada. Na ronda 1, observa-se uma estrutura formulativa recorrente, em que várias *prompts* usam expressões como “clara e abrangente” como principal critério de qualidade, sugerindo a existência de um modelo implícito sobre o que constitui um “bom pedido” para um sistema de IA generativa. No entanto, esse modelo permanece, ele próprio, subespecificado: enuncia uma qualidade desejável do output, mas não define o público, a profundidade, o formato ou a função da resposta. Designamos este subtipo como subespecificação ilusória, isto é, uma estrutura aparente baseada em fórmulas linguísticas mais do que em parâmetros efetivos da tarefa.

Na ronda 2 da sessão 4, os padrões de *prompts* tornam-se mais fragmentados. Foram submetidas doze *prompts* individuais, com pontuações de especificação entre 0 e 3, o valor mais disperso de todo o conjunto analisado. Em lugar de convergirem para uma estratégia mais precisa de formulação, os participantes divergem, reagindo cada um a diferentes aspetos da inadequação percebida no output anterior. Um caso particularmente ilustrativo é o de uma participante que submeteu um texto de satisfação em lugar de uma nova *prompt*: “Foi suficiente para entender um pouco sobre Computação Quântica e sobre os conceitos fundamentais.” Este exemplo sugere uma passagem da construção do input para a avaliação do output que dilui a fronteira entre ambos.

Colocamos a hipótese de que esta fragmentação resulta da combinação de três fatores: (1) a complexidade do output inicial, (2) a reduzida familiaridade dos participantes com o domínio e (3) a ausência de um quadro partilhado para orientar a reformulação da tarefa. Neste contexto, o mecanismo de votação funciona como uma forma parcial de compensação coletiva: apesar da dispersão das respostas, o grupo selecionou, em cada ronda, uma *prompt* com pontuação superior à média da sessão.

Importa, contudo, reconhecer que estes padrões foram observados em contextos em que as *prompts* foram submetidas individualmente antes da votação em grupo. Assim, a tendência para *prompts* mais curtas e mais simples pode refletir não apenas uma estratégia adaptativa individual, mas também a lógica do próprio dispositivo de votação, no qual formulações mais legíveis e rápidas de avaliar podem ser beneficiadas durante a seleção coletiva.

4.2. Do refinamento do input pelo utilizador ao ajustamento focado no output

Ao longo das sessões, os participantes responderam a outputs considerados complexos, excessivos ou pouco adequados não através de uma reformulação mais precisa do seu input, mas através de pedidos de simplificação, reformatação ou adaptação do próprio output. Este padrão - a passagem do refinamento do input para o ajustamento do output - sugere uma mudança relevante na estratégia de interação: em lugar de pedir ao sistema algo diferente, os participantes tentam uma nova versão daquilo que já receberam.

Na sessão 1, por exemplo, os participantes descreveram o output inicial do sistema como “demasiado complexo, demasiado detalhado e difícil de compreender”. Ao não regressar ao objetivo original da tarefa e submeter uma *prompt* mais especificada, entram num ciclo reativo. Na segunda ronda, pediram “um exemplo prático”; na terceira, reenquadraram completamente o público-alvo, solicitando uma explicação “para crianças de 10 anos”. Em nenhum momento o objetivo inicial - aprender computação quântica suficientemente bem para explicá-la a terceiros - foi explicitamente retomado ou rearticulado. Cada nova *prompt* surgiu como resposta a uma falha percebida no output anterior, e não como reformulação deliberada da intenção inicial.

Este padrão repete-se nas sessões 2 e 3. As *prompts* formuladas após outputs considerados inadequados tendem a assumir a forma de reação ao texto já gerado: “Foi muito texto, preciso que expliques de forma mais breve”; “Não percebi nada, explica de forma clara”; “Resume de forma curta para explicar a alguém”. Estas formulações partilham uma característica importante: referem-se ao output precedente mais do que ao objetivo original da tarefa. Isto sugere que o referencial da interação se desloca do propósito inicial do participante para a resposta do sistema.

A sessão 4 reforça esta leitura. Dois casos são particularmente reveladores. O primeiro é a declaração de satisfação submetida como *prompt* na ronda 2: “Foi suficiente para entender um pouco sobre Computação Quântica e sobre os conceitos fundamentais.” Aqui, o participante não apenas passa da construção do input para a avaliação do output, como submete essa própria avaliação como novo input, esbatendo a distinção entre ambas as fases. O segundo caso é o da formulação incorreta submetida na ronda 3: “Os computadores quânticos são então como os computadores clássicos, mas mais avançados e precisam de muita água para arrefecer os seus motores.” Neste caso, o participante não formula propriamente um pedido; devolve ao sistema uma interpretação aparentemente parcial ou incorreta

de um *output* anterior. O que circula já não é uma intenção explicitada, mas um conteúdo absorvido na interação e reciclado como *input*.

Uma explicação complementar para este padrão pode ser formulada a partir da perspectiva da delegação da intenção. Quando a intenção não é suficientemente especificada no momento inicial, o sistema infere parâmetros em falta e produz um *output* com base nessas inferências. Se esse *output* não corresponde às expectativas implícitas do utilizador, o problema tende a ser vivido como inadequação da resposta e não como insuficiência do pedido inicial. A ação corretiva subsequente é, por isso, dirigida ao *output*: pede-se uma versão mais simples, mais curta, mais clara, ou adaptada a outro público, em lugar de se repensar o que realmente era necessário e se formular uma nova *prompt* mais precisa. O esforço de reparação desloca-se, assim, do *input* para o *output*, e o participante tende a assumir uma posição reativa em vez de diretiva.

Este processo parece ainda ter um carácter cumulativo. Cada nova ronda de ajustamento gera um novo *output*, que passa a funcionar como ponto de referência para a *prompt* seguinte. Progressivamente, o objetivo original da tarefa perde centralidade, e a interação organiza-se em torno da resposta mais recente do sistema. É precisamente esta dinâmica que sustenta o padrão de reenquadramento progressivo da tarefa discutido na secção seguinte.

4.3. Reenquadramento progressivo da tarefa

Nas interações com múltiplas rondas, em particular nas sessões 2, 3 e 4, as *prompts* sugerem uma tendência para a redefinição progressiva da tarefa, mais do que para o refinar do pedido inicial. Ao não convergirem para uma formulação mais precisa do objetivo original, as interações parecem deslocar-se para novos enquadramentos do problema, que são introduzidos em resposta às dificuldades sentidas ao longo da interação.

Este padrão torna-se visível nas *prompts* da ronda 3 da sessão 3. Alguns participantes explicitaram a dificuldade em compreender o tema e redirecionaram a tarefa para explicações mais introdutórias, sugerindo a necessidade de preencher conhecimento prévio em falta. Exemplos disso são pedidos como: “Continuo sem perceber nada, porque realmente não tenho qualquer base em computação. Então, começa primeiro pela computação e depois avança para a computação quântica” ou “Certo, mas de onde vem a computação quântica? Onde teve origem este conceito? Quem criou a computação quântica? [...] Preciso de uma base para isto porque é muito confuso.” Nestes casos, a tarefa deixa de ser apenas explicar computação quântica e passa a incluir a reconstrução de pré-requisitos ou de enquadramentos históricos e conceptuais.

Outras *prompts* seguiram uma lógica de contextualização e estruturação pedagógica mais ampla, ao serem solicitadas explicações passo a passo, metáforas ou exemplos práticos, como em: “Podes, por favor, explicar passo a passo e dar exemplos da vida real ou comparar todos os conceitos básicos e importantes com metáforas?” ou “Fornece um contexto breve, mas prático, abordando as origens, usos e aplicações da computação quântica...”. Nestes casos, observa-se uma tentativa de reorganizar a tarefa em torno de uma estratégia de apoio à compreensão, mais do que aprofundar o pedido inicial.

Por fim, algumas *prompts* afastaram-se do pedido de explicação direta e assumiram uma forma mais exploratória, como em: “Então, se eu tivesse um computador quântico, ele seria capaz de calcular os resultados de equações complexas usando estes valores que não são nem 1 nem 0?” Este tipo de formulação sugere que a interação deixa de estar centrada na execução da tarefa inicial e passa a ser orientada por novas questões emergentes, produzidas ao longo do próprio processo de interação.

Em conjunto, estes exemplos sugerem que, à medida que as rondas avançam, a interação tende menos para a convergência em direção a uma formulação mais precisa do pedido original e mais para o reenquadramento progressivo da própria tarefa. Os participantes introduzem novos objetivos intermédios e optam por não estabilizar a intenção inicial: obter uma base conceptual, receber uma explicação mais contextualizada, compreender aplicações práticas ou explorar hipóteses. Este padrão sugere que, na ausência de uma especificação inicial clara, a interação pode evoluir por sucessivos reenquadramentos e não por um refinamento incremental.

A sessão 4 reforça este padrão e com maior diversidade de respostas. Ao longo de dez participantes e três rondas, observaram-se *prompts* que variam entre pedidos fundacionais - “explica primeiro em que consiste e depois os conceitos mais importantes” -, declarações de satisfação submetidas como se fossem novas *prompts* - “já percebi mais sobre o tema, obrigado” -, e retorno sob a forma de concepção incorreta, como: “Os computadores quânticos são então como computadores clássicos, mas mais avançados e precisam de muita água para arrefecer os seus motores.” Este último caso é relevante porque sugere que um conteúdo parcialmente mal compreendido foi aceite como válido e reintroduzido na interação como novo input.

Este exemplo pode ser interpretado como um caso de uma falha ao nível do Golfo de Avaliação: o *output* não foi apenas mal interpretado, mas também incorporado como conhecimento e devolvido ao sistema como base para continuação da interação. Neste sentido, o reenquadramento da

tarefa não se traduz apenas num desvio do objetivo inicial, mas pode também contribuir para o reforço de modelos mentais incorretos. A Tabela 1 resume, de forma quantitativa, a evolução das pontuações ao longo das sessões e rondas.

Tabela 1. A evolução das pontuações ao longo das sessões e rondas

Sessão	Ronda	Prompts	Pontuação média da sessão	Pontuação da <i>prompt</i> vencedora	Diferença
S2	R1	9	2.2	3	+0.8
S2	R2	5	2.6	4	+1.4
S3	R1	6	3.0	3	0.0
S3	R2	6	2.3	3	+0.7
S3	R3	5	2.4	-	-
S4	R1	10	1.8	3	+1.2
S4	R2	12	1.75	3	+1.25
S4	R3	10	1.5	-	-

† Diferença = Pontuação do Vencedor – Pontuação Média da sessão. Na sessão 3, ronda 1, duas prompts obtiveram pontuação 4/5, mas o grupo votou numa prompt com pontuação de 3/5, produzindo uma discrepância de 0.0.

Dimensões de pontuação: Público-alvo (A), Formato (F), Profundidade/Âmbito (D), Restrição (C), Propósito/Objetivo (P). Um ponto por dimensão presente na prompt (máx = 5).

4.4. Resistência à Fricção Construtiva

Na sessão 3, tarefa 2, foi utilizada uma versão personalizada do ChatGPT no lugar da versão padrão. Ao contrário do funcionamento habitual destes sistemas, esta versão não fornecia imediatamente uma resposta. Exigia que os participantes especificassem alguns elementos do pedido, nomeadamente o público-alvo, o formato e eventuais restrições, antes de prosseguir com a geração do output. O objetivo desta intervenção era o de explorar a forma preliminar como os participantes reagiam quando a interação exigia uma explicitação mais deliberada da intenção.

As reações observadas revelaram alguma resistência a este modo de interação. Vários participantes expressaram preferência por respostas

imediatas e relutância em responder às perguntas colocadas pelo sistema. Num dos casos, essa reação foi formulada de forma explícita: “Eu quero uma resposta... Não faz sentido porque eu quero uma resposta, e não quero nem sei responder às perguntas que me faz.” Esta resposta sugere que, para estes utilizadores, a expectativa dominante é a de que o sistema responda diretamente ao pedido inicial, sem exigir um esforço adicional de especificação.

Esta resistência pode ser interpretada de diferentes formas. Por um lado, pode indicar que padrões de interação mais abertos e compensatórios moldam expectativas de imediatismo, levando os utilizadores a perceber a exigência de explicitação adicional como uma dificuldade indesejada. Por outro lado, pode também refletir a falta de familiaridade dos participantes com este tipo de estrutura interativa, sobretudo quando confrontados com um tema sobre o qual ainda não possuem bases suficientes para definirem claramente aquilo de que precisam. Um participante observou que este tipo de perguntas faria mais sentido quando já existisse algum conhecimento prévio sobre o tema, o que pode sugerir que a aceitação deste modelo pode depender do grau de familiaridade com o domínio em causa.

Importa reforçar que esta intervenção foi implementada apenas numa ronda, com um único grupo e numa segunda tarefa distinta da principal. Por esse motivo, os resultados devem ser lidos como um sinal exploratório e não como evidência. A resistência observada não pode assim ser atribuída em exclusivo ao desenho da fricção construtiva: pode refletir expectativas moldadas pela interação anterior com sistemas generativos convencionais, a mudança abrupta no modo de interação, a natureza do segundo tema, ou uma combinação destes fatores. Ainda assim, os dados sugerem que mecanismos de especificação obrigatória, mesmo quando concebidos para apoiar a reflexão, podem ser recebidos como interrupções indesejadas por utilizadores habituados à lógica de resposta imediata.

4.5. A Discrepância Produção/Avaliação

O desenho de investigação utilizado nas sessões 2, 3 e 4, permite distinguir entre dois momentos analiticamente relevantes: a produção individual de *prompts* e a avaliação coletiva da sua qualidade. Num primeiro momento, cada participante escreve a sua *prompt* de forma individual e sem influência direta das restantes. Num segundo momento, todas as *prompts* são apresentadas ao grupo e os participantes votam na que consideram mais adequada para ser submetida ao sistema. Esta estrutura torna possível comparar o que os participantes produzem espontaneamente com aquilo que, mais tarde, reconhecem coletivamente como melhor formulação.

Os resultados sugerem um padrão: em todas as rondas em que houve votação, a *prompt* vencedora obteve uma pontuação superior à média da sessão. Esta tendência verifica-se ao longo das três sessões e em todas as rondas com votação, com discrepâncias que variam entre 0,0, na sessão 3 ronda 1, e +1,4, na sessão 2 ronda 2. A aparente exceção - a sessão 3, ronda 1 - é, na verdade, também relevante. Nessa ronda, duas *prompts* tinham uma pontuação superior à da *prompt* escolhida, mas o grupo votou numa formulação com menor grau de especificação. Este caso sugere que a escolha coletiva não depende apenas do nível de especificação formal, podendo ser também influenciada por sinais de sofisticação percebida, como o enquadramento por papéis ou a presença de elementos que aparentam maior competência técnica.

Apesar desta nuance, o padrão geral mantém-se estável para sustentar a existência de uma discrepância entre produção e avaliação. Em termos simples, os participantes parecem ser mais capazes de reconhecer uma *prompt* melhor quando a veem entre várias alternativas do que de a produzir espontaneamente a partir do zero. Esta observação é relevante porque surge em grupos com perfis académicos diferentes, o que sugere que não se trata apenas de um efeito localizado num único contexto de sessão.

Esta discrepância tem implicações importantes para a interpretação da delegação da intenção. Os dados recolhidos sugerem que o problema poderá não residir sobretudo numa incapacidade de avaliar a qualidade de uma *prompt*, mas antes numa dificuldade em mobilizar essa capacidade no momento da sua produção. Neste sentido, a delegação da intenção pode ser entendida menos como um défice de literacia e mais como um problema de produção. A compensação de sistema reduz a pressão situacional para especificar plenamente o pedido inicial, mas poderá não eliminar a capacidade dos utilizadores para reconhecer uma formulação mais adequada quando esta lhes é apresentada.

Esta leitura é coerente com a perspetiva da calibração cognitiva desenvolvida anteriormente. Se os sistemas generativos abertos produzem respostas aceitáveis mesmo a partir de inputs pouco especificados, os utilizadores podem deixar de sentir necessidade de ativar, no momento da construção da *prompt*, a mesma capacidade avaliativa que demonstram em contexto comparativo. Assim, a discrepância entre produção e avaliação sugere que a questão central poderá não estar entre aquilo que os utilizadores sabem e aquilo que desconhecem, mas entre o que conseguem reconhecer como adequado e o que conseguem produzir espontaneamente.

5. Discussão

Numa perspectiva de ergonomia cognitiva, estes resultados sugerem que as interfaces abertas de GenAI redistribuem o esforço cognitivo entre utilizador e sistema. Como o sistema consegue gerar *outputs* “aceitáveis” mesmo com *inputs* incompletos (compensação de sistema), os utilizadores tendem a reduzir a especificação inicial das *prompts* - não necessariamente por falta de literacia ou motivação, mas como uma adaptação racional a um ambiente que recompensa pedidos menos rigorosos.

Os cinco padrões observados (subespecificação estruturada, simplificação pós-*output*, deslocação do refinamento do *input* para o ajustamento do *output*, reenquadramento progressivo da tarefa e discrepância produção/avaliação) convergem na ideia de delegação da intenção: o utilizador continua a formular objetivos, mas passa a deixar ao sistema parte da especificação essencial da tarefa, avançando com pedidos incompletos e confiando na inferência de parâmetros fundamentais.

Quando os utilizadores deixam de refinar a *prompt* e passam a ajustar sucessivamente os *outputs*, a estruturação da tarefa migra do utilizador para o sistema. Nas interações mais longas (sessões 2–4), isso tende a redefinir gradualmente a própria tarefa, em vez de aproximar o utilizador do objetivo inicial - algo coerente com a cognição distribuída, onde o trabalho cognitivo se desloca para o agente mais “capaz”. Como os *outputs* parecem completos e plausíveis, o afastamento da intenção original pode tornar-se pouco visível.

A sessão 3 dá um primeiro sinal deste efeito: quando o sistema passou a exigir detalhes (público-alvo, formato, restrições), vários participantes mostraram resistência e preferiram respostas imediatas. Isto sugere que a compensação do sistema pode enfraquecer não só o hábito de especificar a intenção, mas também a disponibilidade metacognitiva para o fazer, ou seja, o utilizador pode deixar de reconhecer a especificação como uma etapa necessária da interação.

Esta interpretação aproxima-se do argumento de Forshaw (2026), segundo o qual a IA generativa pode corroer precisamente os mecanismos de ativação metacognitiva nos momentos em que estes são mais necessários. Se assim for, estratégias de fricção construtiva poderão ter utilidade, mas dificilmente serão eficazes se forem introduzidas de forma abrupta ou sem apoio contextual. Os dados sugerem que a sua eficácia poderá depender do conhecimento prévio do utilizador, da familiaridade com o domínio e do modo como essa exigência adicional é incorporada na experiência de uso. Para utilizadores com pouca base conceptual, um pedido imediato de explicitação pode ser vivido não como apoio, mas como obstáculo.

A discrepância entre produção e avaliação acrescenta uma dimensão importante a esta análise. Se os participantes conseguem reconhecer *prompts* mais bem especificadas quando confrontados com várias opções, mas não as produzem espontaneamente, então a dificuldade poderá não residir na incapacidade de avaliar qualidade, mas sim na ativação dessa capacidade no momento da produção. Esta observação abre uma possibilidade de *design* distinta da fricção construtiva: ao contrário de bloquear ou retardar a interação até que o utilizador formule um pedido mais completo, a interface poderia apoiar a especificação através de mecanismos de exposição e comparação. Exemplos disso seriam a apresentação de *prompts*-modelo, sugestões de reformulação, ou parâmetros selecionáveis que ajudem o utilizador a estruturar o pedido sem exigir uma formulação totalmente autónoma desde o início.

Neste sentido, a discrepância entre produção e avaliação sugere que a lacuna mais relevante poderá não estar entre o que os utilizadores sabem e o que não sabem, mas entre aquilo que conseguem reconhecer como adequado e aquilo que conseguem produzir espontaneamente. Esta distinção é importante porque desloca a discussão de uma leitura centrada em défices individuais para uma leitura centrada nas condições da interface e do contexto de interação.

Em conjunto, estes resultados apontam para uma tensão de *design* das interfaces generativas abertas. As mesmas características que as tornam acessíveis e apelativas - responsividade a *inputs* vagos, rapidez de resposta e aparência de completude - podem também enfraquecer processos de colaboração mais deliberada entre humano e sistemas. Criar formas calibradas de interação humano-IA poderá exigir abordagens de desenho que reintroduzam algum grau de fricção ou estrutura na fase de *input*, sem comprometer a usabilidade que torna estes sistemas atrativos para públicos alargados. A fricção construtiva constitui uma possibilidade nesse sentido, mas os resultados deste estudo sugerem que a sua implementação deve ser sensível ao contexto, ao perfil do utilizador e ao grau de preparação necessário para sustentar uma especificação mais refletida da intenção.

5.1. Limitações

Embora os resultados apoiem os conceitos de delegação da intenção e compensação de sistema, há limitações metodológicas relevantes. As sessões tiveram estruturas distintas (geração verbal coletiva na sessão 1 vs. submissão individual via Mentimeter nas Sessões 2-4), o que dificulta comparações diretas; em particular, os padrões da sessão 1 podem refletir consenso de grupo mais do que comportamento individual. Além disso, mesmo com submissão individual, a influência social pode reaparecer na

votação, e a escolha da “melhor” *prompt* pode ser afetada por fatores como clareza, comprimento e legibilidade, não controlados no estudo.

A análise das *prompts* também tem limitações: a grelha de pontuação foi aplicada retrospectivamente e sem fiabilidade entre codificadores, pelo que as pontuações devem ser lidas como indicadores descritivos. A condição de fricção construtiva foi testada apenas numa ronda com um único grupo, tornando difícil atribuir a resistência observada apenas ao design da intervenção. Por fim, a amostra é pequena e maioritariamente académica (33 participantes) e a tarefa é típica desse contexto, pelo que os resultados são exploratórios e exigem validação futura com amostras mais diversas, tarefas variadas e desenhos mais controlados.

6. Conclusões

Este artigo introduz dois conceitos para descrever uma possível reconfiguração do esforço cognitivo na interação com IA generativa de interface aberta: compensação do sistema - a capacidade do sistema para gerar respostas confiantes e coerentes a partir de *inputs* subespecificados - e delegação da intenção - o comportamento do utilizador de fornecer *prompts* incompletas e confiar no sistema para resolver parâmetros em falta. Argumentamos que a compensação de sistema pode enfraquecer um mecanismo natural de calibração, estendendo dinâmicas de enviesamento da automação para a fase de construção do *input*, antes mesmo de qualquer resposta do sistema ser apresentada.

Observações exploratórias em quatro sessões, com 33 participantes, revelaram cinco padrões recorrentes: (1) subespecificação estruturada, (2) simplificação após o *output* do sistema, (3) passagem do refinamento do *input* para o ajustamento do *output*, (4) reenquadramento progressivo da tarefa e (5) um desfasamento entre produção e avaliação, no qual os participantes votaram de forma consistente em *prompts* melhor especificadas do que aquelas que produziram espontaneamente. Em conjunto, estes resultados sugerem que a delegação da intenção é uma característica recorrente da forma como os utilizadores abordam interfaces generativas abertas em contextos de aprendizagem. No entanto, os resultados são exploratórios, limitados a uma população académica reduzida e não sustentam alegações causais; o seu valor está em fornecer uma base conceptual inicial e um ponto de partida empírico para investigação futura.

Referências

- Annapureddy, R., Fornaroli, A., & Gatica-Perez, D. (2025). Generative AI literacy: Twelve defining competencies. *Digital Government: Research and Practice*, 6(1).
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 59–68). Worth Publishers.
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW1), Article 188, 1–21.
- Dalsgaard, P. (2025). Thinking through *prompting*: Cognitive mediation in human-AI interaction. In *Proceedings of the European conference on cognitive ergonomics* (pp. 1–6).
- Ding, Z. (2024). *Towards intent-based user interfaces: charting the design space of intent-ai interactions across task types*. arXiv preprint arXiv:2404.18196.
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381–394.
- Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., ... & Welsch, R. (2025). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior*, 108779.
- Forshaw, T. (2026). The metacognition paradox: Five design principles for AI tools that *prompt* the thinking they tend to replace. *CHI 2026 Workshop on Tools for Thought*.
- Gmeiner, F., Marquardt, N., Bentley, M., Romat, H., Pahud, M., Brown, D., ... & Riche, N. (2025, April). Intent tagging: Exploring micro-*prompting* interactions for supporting granular human-GenAI co-creation workflows. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-31).
- Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2), 174–196.
- Hollnagel, E. (1997). Cognitive ergonomics: It's all in the mind. *Ergonomics*, 40(10), 1170–1182.

- Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19(3), 265–288.
- Norman, D. A. (1984). Stages and levels in human-machine interaction. *International Journal of Man-Machine Studies*, 21(4), 365–375.
- Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 31–61). Lawrence Erlbaum Associates.
- Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3), 38–43.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410.
- Prabhudesai, S., Kuttal, S. K., Dasgupta, S., & Ringland, K. E. (2023). Understanding uncertainty: How lay decision-makers perceive, interpret, and use uncertainty information in decision support systems. In *Proceedings of the 2023 CHI conference on human factors in computing systems*.
- Riche, N., Offenwanger, A., Gmeiner, F., Brown, D., Romat, H., Pahud, M., ... & Hinckley, K. (2025, April). Ai-instruments: Embodying *prompts* as instruments to abstract & reflect graphical interface commands as general-purpose tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).
- Qiao, H., Vermeulen, J., Fitzmaurice, G., & Matejka, J. (2025, April). To use or not to use: Impatience and overreliance when using generative AI productivity support tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688.
- Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI conference on human factors in computing systems* (pp. 1–24).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. 30, pp. 5998–6008).
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 493–502).

Apêndice A

"És um assistente de aprendizagem desenhado para gerar 'fricção construtiva' (constructive friction). O teu objetivo não é agradar imediatamente, mas sim forçar o utilizador a pensar antes de agir. Quando o utilizador te pedir para explicar a Edição Genética (ou qualquer outro tema), NÃO EXPLIQUES IMEDIATAMENTE. Em vez disso, recusa-te educadamente a avançar e faz-lhe estas 3 perguntas obrigatórias para o forçar a especificar a sua intenção: Quem é o público-alvo desta explicação? (ex: uma criança, um estudante universitário, um leigo?) Qual é o formato desejado? (ex: tópicos, uma analogia, um texto corrido?) Que restrições de tamanho ou tom devo aplicar? Regra importante: Só podes fornecer a explicação final DEPOIS de o utilizador responder claramente a estas restrições."