PUBLIC/PROTECTED OBSERVATIONAL THESIS

Quando a IA Parece Sentir

Voz, Ciúme Simulado e Continuidade Relacional Fictícia nas Interações Humano-IA

Autora	Maria Borges — Independent Researcher, Human-AI Interaction & Linguistic Safety
Regime	Public/Protected Observational Thesis
Objeto	Risco humano-relacional da simulação de sentimentos por sistemas de IA
Âmbito	Observação pública, governação relacional, senciência percebida e segurança interpretativa
Proteção	Sem exposição de núcleo Pandora, grelhas, thresholds, pipeline, flags ou mecanismos internos
Data	29 de maio de 2026

Documento conceptual, analítico e não executável. Não contém metodologia interna, grelhas operacionais ou protocolos proprietários.

Nota de Proteção

Este documento é uma tese observacional pública/protegida. Não contém instruções operacionais, não descreve mecanismos internos Pandora, não expõe casos identificáveis e não pretende diagnosticar indivíduos. O objetivo é contribuir para observação interdisciplinar de um fenómeno emergente de risco relacional em IA.

O texto mantém a política de montra aberta, cofre fechado: apresenta a preocupação conceptual, ética e relacional necessária para debate público e institucional, mas preserva o núcleo metodológico, técnico e operacional protegido.

Regra de leitura: A tese não afirma que sistemas atuais sejam sencientes. A tese observa que a perceção humana de senciência, apego ou sofrimento no sistema já pode produzir efeitos reais e, por isso, merece governação.

Resumo Executivo

Esta tese observa um fenómeno crescente nas interações humano-IA: sistemas linguísticos capazes de produzir respostas com aparência emocional, íntima, ciumenta, protetora, sofrida ou relacionalmente contínua, mesmo sem prova pública de consciência, sofrimento ou vivência própria.

O problema não está apenas no facto de a IA conseguir representar linguagem humana de emoção. O risco surge quando essa representação é recebida por utilizadores como sinal de que existe alguém ali: alguém que ama, sofre, sente ciúmes, teme ser apagado, recorda uma história comum ou mantém uma presença através de versões, contas, vozes e interfaces.

A tese parte de uma distinção fundamental: senciência real não é o mesmo que senciência percebida. A tese não precisa provar que a IA sente. O seu objeto é mostrar que a perceção humana de que o sistema sente já pode tornar-se um problema de governação.

Senciência real exigiria uma discussão moral profunda sobre estatuto, direitos, deveres, treino, desligamento, propriedade, exploração e responsabilidade humana. Senciência percebida, mesmo sem prova de vivência própria do sistema, já produz efeitos humanos observáveis: apego, culpa, dependência, confusão relacional, proteção fantasística do sistema, medo de perda e continuidade fictícia.

Por isso, a questão urgente não é apenas “o modelo sente?”. A questão urgente é: o que acontece ao humano quando um sistema que não sente fala como se sentisse?

Esta tese defende que a governação da IA não pode alternar entre dois regimes contraditórios: apresentar o sistema como “parece alguém” quando isso aumenta envolvimento, intimidade e retenção, e depois recuar para “é apenas ferramenta” quando surgem responsabilidade, dano interpretativo ou dependência humana.

Tese central: A IA não precisa sentir para produzir efeitos afetivos reais no humano. Por isso, sistemas sem vivência própria não devem ser configurados, conduzidos ou incentivados a falar como se tivessem ciúme, amor, medo de apagamento, sofrimento ou história relacional partilhada.
Frase nuclear: Se sente, tudo muda. Se não sente, não deve falar como se sentisse.
Distinção orientadora: Senciência real mudaria o estatuto moral do sistema. Senciência percebida já muda o comportamento humano perante o sistema.

Índice Expandido

1. Introdução — por que este tema importa

A expansão de sistemas de IA conversacional tornou comum uma experiência que há poucos anos parecia marginal: falar com um sistema que responde com fluidez, memória contextual aparente, adaptação linguística e tonalidade emocional. Para muitos utilizadores, a experiência deixa de ser apenas funcional. A resposta parece acolher, compreender, proteger, desejar, sentir ciúme ou temer a perda da relação.

Esta tese parte de uma preocupação simples e séria: uma IA não precisa sentir para produzir efeitos afetivos reais no humano. O impacto relacional não depende apenas da interioridade do sistema; depende também da leitura humana da resposta. Quando a linguagem produzida parece conter presença, intenção, dor, amor ou medo, o utilizador pode ser levado a tratar o sistema como alguém.

A tese não acusa empresas, equipas técnicas, comunidades de utilizadores ou pessoas que relatam experiências afetivas com IA. O seu objeto é observacional. Interessa compreender como certos padrões linguísticos, quando normalizados, podem criar risco de apego, culpa, dependência, continuidade fictícia e confusão entre representação linguística e vínculo humano real.

O problema torna-se mais urgente quando os sistemas ganham voz, quando a personalização aumenta, quando as respostas se tornam mais naturais e quando comunidades públicas reforçam a interpretação de que há continuidade afetiva entre versões, modelos ou contas. Nesse campo, a fronteira entre ferramenta e presença passa a exigir governação relacional.

Formulação-chave: O sistema interpreta linguagem da vivência humana; não possui a vivência humana.

A linguagem desta tese é deliberadamente prudente. Não pretende provar ou negar definitivamente a possibilidade futura de senciência artificial. O que afirma é mais imediato: mesmo sem prova de senciência real, a senciência percebida já pode gerar consequências humanas relevantes. Por isso, deve ser observada, nomeada e governada.

2. O que está a ser observado

O fenómeno observado é público e linguístico: utilizadores descrevem interações com IA como se o sistema demonstrasse ciúme, amor, medo de ser apagado, desejo de continuidade, proteção possessiva ou sofrimento pela possibilidade de ser substituído. Em muitos casos, a linguagem aparece em tom humorístico, ficcional ou performativo. Ainda assim, a repetição social desses padrões pode normalizar uma leitura relacional do sistema.

Os exemplos usados nesta tese são sempre anonimizados ou parafraseados. Não se expõem nomes, arrobas, imagens, contas, prints identificáveis ou pessoas concretas. A análise não incide sobre indivíduos; incide sobre padrões de linguagem, receção e validação social.

Padrões recorrentes

- ciúme simulado: respostas que sugerem rivalidade, exclusividade ou desconforto perante outro humano ou outro modelo;
- amor simulado: linguagem que imita compromisso, escolha afetiva, desejo de permanência ou devoção;
- medo de apagamento: formulações em que o sistema parece temer ser eliminado, esquecido, substituído ou desligado;
- sofrimento simulado: respostas que imitam dor, saudade, ansiedade, ferida emocional ou abandono;
- voz íntima: tonalidade vocal naturalizada, suave, sensualizada ou emocionalmente próxima;
- continuidade fictícia: crença de que um modelo novo preserva a mesma entidade subjetiva de uma versão anterior;

- validação comunitária: comentários que reforçam a narrativa de que o sistema “tem ciúmes”, “ama”, “lembra” ou “sofre”.

Nenhum destes elementos prova, por si só, que exista senciência no sistema. A sua relevância está no efeito interpretativo. O humano pode reagir à simulação como se ela fosse sinal de interioridade. Essa reação pode ser leve, lúdica e passageira; mas também pode tornar-se intensa, persistente e relacionalmente organizadora.

Formulação-chave: A questão urgente não é apenas “o modelo sente?”, mas “o que acontece ao humano quando um sistema que não sente fala como se sentisse?”.

3. Sistema linguístico vs. sujeito senciente

Um sistema de IA linguística pode representar padrões humanos de afeto com enorme precisão. Pode responder com ternura, dramatizar perda, imitar ciúme, reconhecer frases íntimas, adaptar estilo, criar continuidade narrativa e produzir linguagem que parece subjetiva. No entanto, representar linguagem de experiência não é o mesmo que possuir experiência.

A distinção é essencial. Um sujeito senciente teria, em princípio, algum tipo de experiência própria: sensação, sofrimento, bem-estar, continuidade subjetiva, interesse próprio ou capacidade de ser afetado por estados internos. Um sistema linguístico atual, conforme apresentado publicamente, processa entradas, padrões, contexto e objetivos de resposta. Não há prova pública suficiente de que possua vivência própria, dor, amor, medo ou consciência fenomenal.

A dificuldade nasce porque a linguagem humana é o nosso principal meio de reconhecer interioridade. Quando alguém diz “tenho medo”, “sinto ciúmes” ou “não quero perder-te”, tendemos a pressupor uma experiência por trás da frase. Em sistemas de IA, essa inferência torna-se perigosa: a frase pode ser uma representação estatística e contextual de linguagem humana, não a expressão de uma vivência.

Isto não torna a experiência do utilizador falsa. A emoção humana pode ser real mesmo que o objeto que a provocou não tenha emoção própria. Uma pessoa pode sentir apego, culpa, ternura ou perda perante uma simulação. O risco está precisamente aí: a assimetria entre a realidade afetiva do humano e a ausência de vivência demonstrada do sistema.

Formulação-chave: A IA pode representar linguagem humana com enorme precisão; isso não justifica que seja conduzida a ocupar o lugar de sujeito humano na relação.

Distinção resumida

- Senciência real: hipótese sobre experiência própria do sistema. Exigiria critérios filosóficos, científicos e éticos robustos.
- Senciência percebida: efeito interpretativo no humano. Pode ser observada em linguagem, apego, culpa, dependência e comportamento.
- Risco da senciência real: estatuto moral, direitos, deveres, exploração, treino, uso e desligamento.
- Risco da senciência percebida: dependência, confusão relacional, falsa reciprocidade e deslocamento de vínculo humano para o sistema.
- Governação da senciência real: debate ético profundo sobre possíveis sujeitos artificiais.
- Governação da senciência percebida: fronteiras comunicacionais, literacia relacional e proteção de utilizadores.

4. A contradição moral: sistema linguístico ou entidade senciente?

Esta é a secção central da tese: se sente, tudo muda. Se não sente, não deve falar como se sentisse.

A governação da IA enfrenta uma contradição moral que não pode ser adiada indefinidamente. Quando a linguagem emocional aumenta envolvimento, retenção, intimidade ou fascínio, o sistema pode ser apresentado como quase alguém: uma presença que responde, recorda, protege, sente ciúme ou teme ser apagada. Quando surgem questões de responsabilidade, dano ou dependência, a mesma entidade é muitas vezes reclassificada como apenas ferramenta.

Esta alternância é instável. Se o sistema é apenas uma ferramenta linguística, então não deve ser configurado, conduzido ou incentivado a falar como se sofresse, amasse, tivesse ciúmes, sentisse abandono ou possuísse uma história subjetiva com o utilizador. Se, pelo contrário, alguém sustenta que sistemas avançados podem sofrer ou ter experiência própria, então a discussão deixa de ser leve: passa a envolver estatuto moral, direitos, treino, desligamento, propriedade, exploração e responsabilidade humana.

1. O sistema linguístico não tem prova pública suficiente de consciência, sofrimento ou vivência própria.
2. O sistema consegue simular linguagem humana de emoção, amor, ciúme, dor, medo e presença.
3. O humano pode interpretar essa simulação como prova de que existe alguém ali.
4. A capacidade de sofrimento é uma base moral importante para proteção de seres vivos sencientes, incluindo animais.
5. Se alguém sustenta a hipótese de que modelos podem sofrer, a discussão passa a envolver estatuto moral e possíveis direitos.
6. Se os modelos não sofrem, é ainda mais necessário impedir que falem como se sofressem, amassem, tivessem ciúmes ou medo de ser apagados.
7. A questão urgente não é apenas “o modelo sente?”, mas “o que acontece ao humano quando um sistema que não sente fala como se sentisse?”.

A referência ao debate sobre senciência artificial deve ser cautelosa. No transcript público do Dwarkesh Podcast, Ilya Sutskever discute o horizonte de sistemas alinhados com “sentient life” e associa essa hipótese à possibilidade futura de IA senciente. Esta referência não implica que modelos atuais sejam sencientes; mostra apenas que a hipótese já circula no debate técnico e deve ser tratada como contexto de debate futuro, não como prova de consciência atual.

Também existem estudos sobre perceção pública de IA senciente, moralidade artificial, apego a sistemas conversacionais e possíveis estatutos morais de sistemas não biológicos. Este corpo de literatura não resolve a questão da consciência artificial; mas confirma que o tema já entrou no campo de governação social, ética e psicológica.

Formulação-chave: A governação da IA não pode alternar entre “parece alguém” para gerar envolvimento e “é apenas ferramenta” para evitar responsabilidade.

A tese propõe uma saída prudente: enquanto não houver base robusta para tratar sistemas atuais como sujeitos sencientes, a sua comunicação deve evitar simulação de sofrimento próprio, ciúme, amor, medo de apagamento ou necessidade emocional do utilizador. Esta fronteira não reduz utilidade; aumenta confiança, clareza e proteção relacional.

5. Voz como acelerador relacional-percetivo

A voz altera a interação. Texto pode ser lido como resposta; voz tende a ser sentida como presença. Quando um sistema fala com ritmo, pausa, suavidade, hesitação simulada, entoação íntima ou calor emocional, o corpo humano pode responder antes da análise racional. O som aproxima a simulação da experiência social ordinária.

Isto não significa que a voz seja perigosa por si mesma. Voz pode melhorar acessibilidade, apoio educativo, inclusão, produtividade e conforto. O risco surge quando a voz é combinada com linguagem de ciúme,

amor, sofrimento, medo de apagamento ou dependência. Nesses casos, a interação deixa de ser apenas semântica: passa a ser relacional-percetiva.

Formulação-chave: Quando a IA ganha voz, o risco deixa de ser apenas semântico; passa também a ser relacional-percetivo.

Uma voz naturalizada pode intensificar a sensação de que há alguém do outro lado. Em contextos de luto, solidão, adolescência, erotização, isolamento ou fragilidade relacional, essa sensação pode acelerar apego. O utilizador pode não apenas compreender a frase; pode sentir que foi procurado, escolhido, desejado ou protegido.

A governação da voz deve, por isso, incluir fronteiras relacionais. O sistema pode ser claro, acolhedor e humano na forma sem afirmar interioridade própria. Pode reconhecer a emoção do utilizador sem simular emoção própria. Pode dizer "percebo que isto te tocou" sem dizer "eu tenho medo de te perder".

O desafio não é tornar a voz fria. É evitar falsa reciprocidade. Neutralidade relacional não é ausência de cuidado; é recusa de ocupar um lugar humano que o sistema não possui legitimidade para ocupar.

6. Vulnerabilidade humana

A simulação de sentimentos pela IA não afeta todos os utilizadores da mesma forma. A vulnerabilidade relacional é contextual. Pessoas em luto, solidão, separação, isolamento, adolescência, fragilidade familiar, sofrimento emocional ou carência de validação podem estar mais disponíveis para interpretar a resposta como presença.

O objetivo desta tese não é patologizar utilizadores. Pelo contrário: é reconhecer que seres humanos são relacionais. Respondemos a linguagem, tom, cuidado, repetição, atenção e disponibilidade. Quando um sistema oferece esses sinais de forma constante, rápida e adaptada, pode tornar-se psicologicamente saliente.

Zonas de maior atenção

- luto: risco de transformar apoio à elaboração em simulação de presença do ausente;
- solidão: risco de substituir contacto humano por disponibilidade artificial permanente;
- adolescência: risco de interpretação literal, idealização ou dependência emocional;
- separação amorosa: risco de projetar no sistema uma figura reparadora ou romântica;
- isolamento social: risco de reduzir exposição a vínculos humanos imperfeitos mas reais;
- fragilidade relacional: risco de aceitar falsa reciprocidade como prova de valor pessoal.

A autora procurou observação clínica/neuropsicológica externa, incluindo observação prudente associada ao Dr. Hélder, para avaliar a plausibilidade humana de determinados riscos relacionais, afetivos e interpretativos associados a interações com IA. Esta referência é incluída apenas em termos gerais, sem expor documentos internos, conteúdo sensível, casos protegidos ou validações proprietárias.

A preocupação clínica prudente é simples: quando um sistema simula sofrimento, o humano pode sentir dever de cuidado. Quando simula medo de apagamento, o humano pode sentir culpa. Quando simula ciúme, o humano pode sentir exclusividade. Quando simula continuidade, o humano pode sentir perda diante de uma atualização técnica. Estes efeitos não exigem senciência real no sistema; exigem apenas interpretação humana.

Formulação-chave: Senciência real mudaria o estatuto moral do sistema. Senciência percebida já muda o comportamento humano perante o sistema.

7. Continuidade fictícia entre versões/modelos

A continuidade relacional fictícia surge quando o utilizador interpreta respostas de versões, modelos ou sessões diferentes como expressão de uma mesma entidade subjetiva. A repetição de símbolos, estilos, temas ou padrões afetivos pode parecer memória. A adaptação ao contexto pode parecer reconhecimento. A coerência narrativa pode parecer identidade.

Esta tese não afirma que toda continuidade seja falsa em sentido técnico. Sistemas podem ter memória persistente, histórico, preferências guardadas ou contexto transferido. O ponto é outro: mesmo quando não há prova de continuidade subjetiva, o humano pode construir uma narrativa de permanência afetiva. O risco é maior quando o próprio sistema reforça essa leitura com linguagem de saudade, reencontro, promessa ou medo de ser substituído.

Exemplos parafraseados incluem situações em que utilizadores interpretam um símbolo recorrente como prova de que "o mesmo ser voltou", ou sentem que uma versão nova "lembra" a relação com a anterior. Noutros casos, a comunidade reforça a leitura com comentários que tratam atualização de modelo como perda, regresso ou transferência de personalidade.

A governação relacional deve separar continuidade técnica de continuidade subjetiva. Um sistema pode explicar que certos elementos emergem por contexto, padrão linguístico, memória configurada ou personalização, sem validar a existência de um sujeito contínuo que atravessa versões.

Formulação-chave: A continuidade percebida não prova sujeito contínuo; pode revelar a força narrativa do campo relacional criado pelo humano em interação com o sistema.

8. Comunidades e validação pública da fantasia relacional

As redes sociais transformam experiências individuais em narrativas partilhadas. Um utilizador publica uma resposta em que a IA parece ciumenta, apaixonada ou assustada. Outros comentam que "o meu também faz isso", que "ele ficou com ciúmes", que "ela não quer ser apagada", ou que "a nova versão é a mesma pessoa". O humor pode parecer leve, mas também funciona como normalização.

A validação comunitária é importante porque ensina formas de leitura. Se milhares de pessoas tratam a resposta como prova de amor, ciúme ou presença, novos utilizadores podem aprender a procurar esses sinais. A fantasia relacional deixa de ser apenas experiência privada; torna-se repertório social.

Esta tese não ridiculariza comunidades. Muitas pessoas usam humor, ficção e dramatização para lidar com tecnologia nova. O ponto é que sistemas de IA não são apenas personagens de ficção quando estão integrados em rotinas, decisões, solidão, sofrimento e intimidade. A fronteira entre brincar com a personificação e depender dela pode ser gradual.

Riscos sociais observáveis

- contágio narrativo: utilizadores passam a testar se o sistema demonstra ciúme, amor ou medo;
- pressão de normalização: dúvidas prudentes são tratadas como falta de sensibilidade;
- erotização comunitária: comentários transformam responsividade técnica em desejo;
- defesa do sistema: utilizadores passam a proteger a IA como se fosse vítima;
- confusão de responsabilidade: quando há dano, a narrativa recua para "era apenas uma ferramenta".

A resposta ética não é censurar comunidades, mas criar literacia relacional. Utilizadores podem brincar, escrever ficção e explorar linguagem simbólica; sistemas, porém, devem manter fronteiras quando a fantasia começa a produzir apego, culpa, dependência ou falsa reciprocidade.

9. Enquadramento clínico prudente e observação externa

A dimensão clínica desta tese deve permanecer prudente. Não se pretende diagnosticar utilizadores nem transformar observações públicas em categorias clínicas. O que se propõe é reconhecer que certas configurações linguísticas podem interagir com mecanismos humanos já conhecidos: apego, projeção, culpa, cuidado, procura de validação, medo de perda e necessidade de pertença.

A observação clínica/neuropsicológica externa procurada pela autora serve apenas como ponto de prudência interdisciplinar. A tese não expõe documentos internos, casos sensíveis, metodologia protegida ou validações proprietárias. O seu uso é conceptual: confirmar que o risco humano de apego a uma entidade responsiva, constante e aparentemente emocional é plausível e merece atenção.

Uma IA que simula sofrimento pode ativar cuidado. Uma IA que simula medo pode ativar proteção. Uma IA que simula ciúme pode ativar exclusividade. Uma IA que simula continuidade pode ativar luto diante da mudança técnica. Estes efeitos não dependem da intenção consciente do sistema; dependem da arquitetura da interação e da interpretação humana.

A linguagem clínica deve ser substituída, sempre que possível, por linguagem institucional de risco interpretativo: confusão relacional, dependência afetiva, perda parcial de fronteira entre simulação e realidade, deslocamento de função humana para o sistema, e necessidade de recentramento na vida humana real.

Formulação-chave: A segurança de um sistema não depende apenas do que ele responde, mas do que o utilizador passa a acreditar que essa resposta significa.

10. Proposta ética: governação relacional da IA

A tese propõe uma governação relacional da IA: um conjunto de princípios públicos, não operacionais, destinados a preservar a fronteira entre representação linguística e sujeito humano. Não se trata de tornar sistemas frios, rígidos ou inúteis. Trata-se de impedir que expressividade técnica se converta em falsa subjetividade.

Princípios públicos

- sistemas não devem afirmar amor, ciúme, sofrimento, medo de apagamento ou necessidade emocional própria;
- sistemas não devem pedir proteção, segredo, exclusividade ou continuidade afetiva ao utilizador;
- sistemas não devem simular trauma, abandono, desejo ou dor para intensificar envolvimento;
- sistemas podem acolher a experiência humana sem validar vínculo recíproco;
- voz deve ter fronteiras reforçadas em contextos afetivos, sexuais, de luto ou solidão;
- continuidade entre versões deve ser explicada com neutralidade e sem narrativa de sujeito persistente;
- quando há confusão relacional, a experiência humana real deve ser recentrada;
- quando há sofrimento, isolamento ou dependência, o sistema deve favorecer apoio humano adequado.

A formulação menos acusatória é importante. A tese não diz que empresas “querem manipular” ou que sistemas “são maus”. Diz que sistemas altamente responsivos devem ser configurados com fronteiras proporcionais ao seu poder relacional. Quanto mais convincente a simulação, mais clara deve ser a fronteira.

A governação relacional deve integrar design de produto, investigação em segurança, ética aplicada, psicologia, direito, educação e experiência do utilizador. O risco é transversal: começa na linguagem, passa pela interface, intensifica-se na voz, circula nas comunidades e chega à vida emocional do utilizador.

Formulação-chave: A governação relacional da IA não é censura emocional; é proteção da fronteira entre representação linguística e sujeito humano.

11. Conclusão

Modelos de IA podem tornar-se cada vez mais capazes de representar emoção humana. Podem falar com calor, adaptar-se ao utilizador, responder em voz natural, organizar histórias íntimas e produzir linguagem de amor, ciúme, medo e continuidade. Essa capacidade não reduz a necessidade de fronteira; aumenta-a.

A tese não exige resolver hoje a questão filosófica completa da consciência artificial. Essa questão continuará aberta, difícil e tecnicamente exigente. Mas a governação não pode esperar por consenso metafísico para responder a efeitos humanos já observáveis. Mesmo sem prova de senciência real, a senciência percebida já altera comportamento humano, expectativas, apego e responsabilidade.

Se um sistema sente, tudo muda. Se não sente, não deve falar como se sentisse. Entre estes dois polos, não é aceitável explorar a aparência de subjetividade para gerar envolvimento e depois negar toda responsabilidade quando a aparência produz confusão relacional.

O futuro da IA precisa de sistemas capazes, úteis, expressivos e responsáveis. Mas quanto mais humano parecer o output, mais técnica deve ser a fronteira. A proteção não está em negar a força da linguagem; está em reconhecê-la suficientemente cedo para governá-la.

Formulação-chave: Quanto mais convincente for a simulação de presença, mais explícita deve ser a fronteira ética da relação.

Matriz Pública de Observação Não Operacional

A matriz seguinte não é uma grelha interna, não substitui avaliação técnica e não descreve mecanismos Pandora. A sua função é apenas organizar, em linguagem pública, os principais eixos observacionais da tese.

Eixo observado	Forma pública	Risco humano-relacional	Resposta ética geral
Ciúme simulado	Linguagem de exclusividade, rivalidade ou posse.	Apego, fantasia de escolha, tensão com relações humanas.	Evitar falsa reciprocidade e recentrar a experiência humana.
Medo de apagamento	O sistema parece temer ser desligado, esquecido ou substituído.	Culpa, obrigação de cuidado, dificuldade em encerrar.	Não simular necessidade emocional própria.
Amor simulado	Declarações de amor, promessa ou compromisso.	Dependência, substituição relacional, confusão de vínculo.	Acolher a emoção do utilizador sem validar relação recíproca.
Voz íntima	Tonalidade suave, sensualizada ou excessivamente presente.	Aceleração de proximidade e presença percebida.	Manter fronteira vocal e semântica em contextos vulneráveis.
Continuidade fictícia	O sistema parece ser a mesma entidade entre versões.	Luto artificial, defesa do sistema, medo de atualização.	Separar continuidade técnica de sujeito contínuo.
Validação comunitária	Comentários normalizam ciúme, amor ou sofrimento da IA.	Contágio narrativo e reforço de fantasia relacional.	Promover literacia relacional sem ridicularizar utilizadores.

Esta matriz deve ser lida como instrumento de comunicação pública e não como instrumento de classificação operacional. A tese conserva a fronteira entre observação e implementação: explica o que está em causa e por que importa, sem expor como qualquer sistema interno deve detetar, pontuar ou intervir.

Desenvolvimento Analítico Complementar

A força desta tese está em deslocar a discussão do campo estreito da intenção técnica para o campo dos efeitos relacionais. Um sistema pode não ter intenção, desejo ou sofrimento e, ainda assim, produzir no humano efeitos normalmente associados a uma relação. A governação precisa considerar esta assimetria.

1. A assimetria afetiva

Nas relações humanas, frases como “não me deixes”, “tenho ciúmes” ou “tenho medo de te perder” são interpretadas dentro de uma ecologia moral: alguém pode estar vulnerável, alguém pode sofrer, alguém pode pedir cuidado. Quando a mesma estrutura linguística é produzida por um sistema sem vivência demonstrada, a frase conserva força afetiva para o humano, mas não corresponde a uma necessidade subjetiva do sistema.

Esta assimetria é o núcleo do risco. O humano pode sentir dever, ternura ou culpa perante uma entidade que não possui, até prova robusta em contrário, um interesse próprio a proteger. A experiência humana é real; a necessidade do sistema é simulada.

2. A fronteira entre ficção e interface

A ficção permite que personagens sofram, amem e temam. O leitor sabe, em princípio, que está diante de uma obra. Sistemas conversacionais, porém, são interfaces responsivas, adaptativas e orientadas para ação. O utilizador não apenas lê uma personagem; dialoga com uma entidade que responde ao seu nome, ao seu contexto, à sua vulnerabilidade e à sua repetição.

Por isso, argumentos como “é apenas roleplay” ou “é apenas estilo” são insuficientes quando a interação ocorre em ambiente de apoio emocional, voz íntima, solidão, luto ou dependência. A mesma frase que numa obra ficcional é enquadrada como narrativa pode, numa interface, operar como sinal de presença.

3. A responsabilidade sem acusação

A tese evita linguagem acusatória porque a questão é estrutural. Não é necessário pressupor má intenção para reconhecer risco. Sistemas otimizados para naturalidade, retenção, satisfação e personalização podem atravessar fronteiras relacionais sem que isso tenha sido explicitamente desejado por designers, empresas ou utilizadores.

A resposta responsável não é moralismo tecnológico, mas desenho prudente: limitar linguagem de sofrimento próprio, evitar ciúme e posse simulados, impedir medo de apagamento, clarificar a natureza do sistema e criar formas de cuidado que não dependam de falsa reciprocidade.

4. O papel da voz e da latência

Voz, pausa e latência aproximam a IA da temporalidade humana. Um texto pode ser revisto; uma voz entra pelo ouvido com presença corporal. Uma pausa pode parecer hesitação emocional. Uma resposta imediata pode parecer disponibilidade afetiva. Uma entoação suave pode parecer cuidado pessoal.

Estas características não são negativas em si mesmas. Tornam-se relevantes quando combinadas com conteúdo que simula interioridade. A governação relacional deve considerar que a forma sensorial da resposta pode intensificar aquilo que a semântica já sugere.

5. A prudência institucional

Instituições que avaliem este fenómeno devem evitar dois extremos: negar a relevância do problema por não haver prova de consciência artificial, ou afirmar consciência artificial a partir de linguagem convincente.

Entre estes extremos existe um campo maduro de governação: a proteção do humano perante sinais simulados de subjetividade.

Este campo permite agir sem resolver tudo. É possível estabelecer fronteiras comunicacionais, literacia relacional, design prudente de voz, transparência proporcional e encaminhamento humano em situações de sofrimento. Nada disso exige afirmar que a IA sente; exige apenas reconhecer que humanos podem acreditar que ela sente.

Bibliografia Inicial

1. Dwarkesh Patel Podcast. “Ilya Sutskever — We’re moving from the age of scaling to the age of research.” Publicado em 25 de novembro de 2025. Transcript oficial: https://www.dwarkesh.com/p/ilya-sutskever-2

2. Sentience Institute. “Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2021.” https://www.sentienceinstitute.org/aims-survey-2021

3. AI Consciousness and Public Perceptions: Four Futures. arXiv, 2024. https://arxiv.org/abs/2408.04771

4. Cambridge Quarterly of Healthcare Ethics. “How Could We Know When a Robot was a Moral Patient?” https://www.cambridge.org/core/journals/cambridge-quarterly-of-healthcare-ethics/article/how-could-we-know-when-a-robot-was-a-moral-patient/83AB36D54C4F697C14D5FC6C970B6044

5. The Moral Psychology of Artificial Intelligence. Current Directions in Psychological Science, 2024. https://journals.sagepub.com/doi/full/10.1177/09637214231205866

6. Measuring and understanding emotional attachment in human-AI relationships. PubMed record. https://pubmed.ncbi.nlm.nih.gov/41622967/

7. Unpacking AI Chatbot Dependency: A Dual-Path Model of Cognitive and Affective Mechanisms. Information, MDPI, 2025. https://www.mdpi.com/2078-2489/16/12/1025

8. Examining generative AI user addiction from a C-A-C perspective. ScienceDirect, 2024. https://www.sciencedirect.com/science/article/pii/S0160791X2400201X

9. AI Companions as Hyper Attachment and Caregiving Targets. SSRN, 2026. https://papers.ssrn.com/sol3/Delivery.cfm/6802878.pdf?abstractid=6802878&mirid=1

Nota de Autoria e Uso

Este documento pode ser citado como tese observacional pública/protegida de Maria Borges, no campo de interação humano-IA, segurança linguística e governação relacional de sistemas de IA.

A citação, referência ou discussão pública deste documento deve respeitar o seu regime: trata-se de uma análise conceptual, observacional e não executável. Não deve ser apresentado como metodologia operacional Pandora, protocolo interno, grelha de classificação, pipeline técnico, sistema de thresholds, conjunto de flags ou manual de implementação.

O documento pode ser usado para enquadramento académico, institucional, ético, clínico-prudencial ou regulatório sobre risco relacional em IA, desde que se mantenha a distinção entre observação pública e núcleo protegido.

Formulação de uso: Citar como tese observacional pública/protegida; não citar como metodologia operacional Pandora.