ARQUITETURAS ADVERSÁRIAS NA NEUROCOGNIÇÃO DO PROPÓSITO E PERCEPÇÃO

Hélio Craveiro Pessoa Júnior[1]

[1]Universidade de Brasília, Brasília, Brasil (pessoajr@gmail.com)

Resumo: Este relatório analisa percepção, sono e propósito sob a ótica de arquiteturas adversárias inspiradas nas GANs. Investiga-se a Teoria da Ativação Defensiva no sono REM, a Hipótese do Cérebro Sobreajustado nos sonhos e a neurobiologia do propósito como atrator dinâmico (alostase). Discutem-se implicações clínicas (coma, BCIs), ameaças de guerra cognitiva e a governança baseada nos loops Eyla e PAGRL, fornecendo um framework integrado e rigorosamente referenciado.

Palavras-chave: Cérebro Preditivo; Redes Adversárias Generativas; Sono REM; Propósito; Neurotecnologia.

INTRODUÇÃO

A literatura recente em neurociência computacional e inteligência artificial reforça uma inflexão importante: a mente deixa de ser tratada apenas como receptora de sinais e passa a ser modelada como sistema de predição ativa (SHAW et al., 2026). Neste enquadramento, a percepção consciente e a construção de sentido ou propósito passam a ser descritas como efeitos de um equilíbrio dinâmico entre forças internas concorrentes .). O paradigma da codificação preditiva e adversarial sugere, em chave heurística e parcialmente testável, que o cérebro opera sob uma lógica análoga à das Redes Adversárias Generativas (GANs), em que um componente "Gerador" cria simulações da realidade enquanto um "Discriminador" as avalia quanto à plausibilidade e coerência. Este estudo detalha contribuições relevantes para a área entre 2024 e 2026, com ênfase na defesa territorial do córtex, na mitigação do sobreajuste cognitivo por meio do sono e na emergência de arquiteturas de governança neurocognitiva em resposta às ameaças de manipulação algorítmica.

O argumento central deste relatório é que percepção, sonho e propósito podem ser analisados como produtos de uma ecologia adversarial do cérebro, na qual geração interna de hipóteses e discriminação de erros mantêm uma tensão produtiva (SHAW et al., 2026; EAGLEMAN; VAUGHN, 2021; PRINCE; RICHARDS, 2021). Essa tese é desdobrada em cinco frentes: cérebro preditivo, defesa territorial do córtex visual, função anti-sobreajuste dos sonhos, papel neurobiológico do propósito e implicações clínicas e políticas das arquiteturas adversárias contemporâneas (SHAW et al., 2026; EAGLEMAN; VAUGHN, 2021; PRINCE; RICHARDS, 2021; PUBMED, 2026).

O texto também sustenta uma tese normativa: quanto mais a cognição humana depende de mecanismos plásticos, preditivos e afetivamente orientados, maior é a necessidade de mecanismos explícitos de governança, auditoria inferencial e higiene epistemológica diante de sistemas digitais capazes de modular atenção, crença e decisão (COGNITIVE NEUROSCIENCE SOCIETY, 2026; THE TELEONOME..., 2026; ĆOSIĆ et al., 2026; BANDARA et al., 2026).

A migração do modelo cerebral de processamento de sinais para o de predição ativa representa uma mudança de paradigma relevante na década. A evidência acumulada entre 2024 e 2025 indica que o cérebro não aguarda o input sensorial para formar uma imagem do mundo; em vez disso, ele gera constantemente hipóteses top-down que são testadas contra o erro de predição bottom-up (SHAW et al., 2026). Dentro dessa mecânica, a analogia com as Generative Adversarial Networks (GANs) deixou de ser meramente metafórica para se tornar uma proposta biológica testável. Nas GANs, dois jogadores — o Gerador e o Discriminador — travam um jogo de soma zero em que a precisão de um depende do desafio imposto pelo outro (ZHAO et al., 2026).

No contexto neurobiológico, pesquisas recentes sugerem uma especialização hemisférica que mimetiza essa estrutura: em chave heurística, o hemisfério direito, com sua maior conectividade inter-regional e capacidade de síntese, funcionaria

como o Gerador de narrativas, simulações de futuro e construções de identidade (BUZEA; AGOP, 2025). Em contrapartida, também em chave heurística, o hemisfério esquerdo, voltado para a análise, lógica e verificação de regras, atuaria como o Discriminador, confrontando essas narrativas com o feedback do ambiente e a coerência lógica. A saúde mental, sob esta ótica, depende da integridade dessa "tensão" adversarial (SHAW et al., 2026). Quando o Gerador domina sem a devida supervisão do Discriminador, observam-se fenômenos de psicose e alucinações; quando o Discriminador é excessivamente rígido, o resultado é a rigidez cognitiva e a perda de propósito.

Tabela 1. Correspondência entre os atores adversariais neurocognitivos e os componentes das GANs

Ator Adversarial	Função Neurocognitiva Proposta	Correlato em Inteligência Artificial
Gerador Interno	Produção de simulações de futuro, sonhos e narrativas de identidade no hemisfério direito	Criação de amostras sintéticas a partir de vetores de ruído latente (ZHAO et al., 2026)
Discriminador Realista	Avaliação da plausibilidade sensorial e social; filtragem de erros no hemisfério esquerdo	Classificação de dados como "reais" (origem externa) ou "falsos" (origem interna) (ZHAO et al., 2026)
Erro de Predição	Tensão entre desejo, projeto de si e realidade física/social; motor da adaptação humana	Função de custo Minimax que orienta o treinamento do modelo adversarial (VARUGHESE, s.d.)
Estabilidade Dinâmica	Percepção consciente equilibrada; senso de propósito resiliente e adaptável	Equilíbrio de Nash, no qual nem o gerador nem o discriminador conseguem melhorar sem cooperação mútua (ZHAO et al., 2026)

Este equilíbrio adversarial é importante para a construção do propósito. O propósito não é visto como um dado estático ou uma herança genética imutável, mas como um "atrator" em um sistema complexo que se auto-otimiza por meio do conflito entre o que o indivíduo deseja ser e o feedback que o mundo devolve, conforme a Figura 1. Estudos de 2025 demonstram que indivíduos com maior clareza de propósito exibem uma flexibilidade cognitiva superior, pois seu "Discriminador" interno é treinado para integrar contradições sem colapsar a estrutura de identidade.

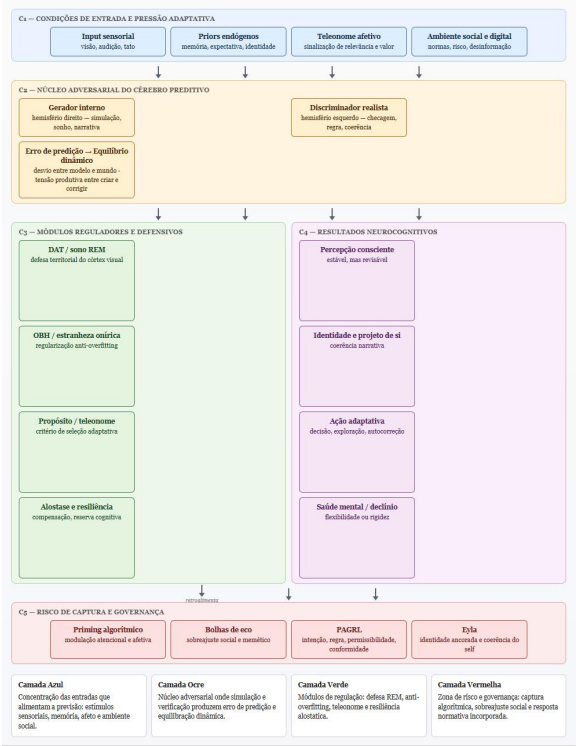

Figura 1. Dinâmica adversarial entre o Gerador interno de hipóteses e o Discriminador de plausibilidade, com o erro de predição convergindo para o propósito como atrator estável.

MATERIAL E MÉTODOS

Este relatório combina artigos revisados por pares, preprints, textos de divulgação científica e materiais institucionais. A leitura proposta é estratificada: referências primárias ocupam o núcleo dos mecanismos discutidos, enquanto fontes de divulgação e repositórios sociais aparecem apenas como apoio contextual, em coerência com a

hierarquia argumentativa adotada nas seções analíticas centrais (SHAW et al., 2026; EAGLEMAN; VAUGHN, 2021). Ao longo do texto, verbos como "sugere", "indica" ou "propõe" preservam a distinção entre evidência consolidada, inferência plausível e extrapolação heurística.

O critério de inclusão prioriza publicações entre 2024 e 2026 quando elas atualizam aplicações clínicas, governança algorítmica ou neurotecnologia, mas mantém trabalhos anteriores quando estes funcionam como referência fundacional, como a DAT e a OBH (EAGLEMAN; VAUGHN, 2021; PRINCE; RICHARDS, 2021). As evidências consolidadas são apresentadas como mecanismos com suporte empírico direto; as inferências plausíveis conectam achados de campos próximos; e as extrapolações heurísticas são marcadas como modelos interpretativos, não como equivalências causais ou anatômicas literais (SHAW et al., 2026). Esse procedimento reduz o risco de transformar analogias produtivas em afirmações causais excessivas.

Como apoio de consistência interna, o documento pode ser auditado pelo script validator_logico.py, uma ferramenta local em Python estruturada como Toolkit do agno, voltada para detectar artefatos textuais.

RESULTADOS E DISCUSSÃO

Teoria da Ativação Defensiva: Territorialidade e a Guerra Noturna Sináptica

Um pilar central da tese adversarial moderna reside na Teoria da Ativação Defensiva (DAT), expandida significativamente pelos trabalhos de David Eagleman e Don Vaughn da Universidade de Stanford entre 2024 e 2026 (EAGLEMAN; VAUGHN, 2021). A DAT parte do princípio de que o cérebro humano é um sistema "livewired" (fiação viva), em que os neurônios estão em uma luta constante por "espaço sináptico" (PORTIER, 2026). Diferente do hardware de um computador, o cérebro reorganiza seus circuitos em tempo real com base no uso (EAGLEMAN; VAUGHN, 2021). Áreas não utilizadas não ficam ociosas; elas são rapidamente "sequestradas" ou "colonizadas" por sentidos vizinhos que continuam recebendo estímulos (EAGLEMAN; VAUGHN, 2021).

A Fragilidade do Sistema Visual e o Sono REM

O sistema visual enfrenta um problema evolutivo único: devido à rotação do planeta, os organismos foram submetidos a ciclos de doze horas de escuridão durante a maior parte da história biológica (EAGLEMAN; VAUGHN, 2021). Enquanto o tato, a audição e o olfato permanecem ativos no escuro, permitindo detectar um predador ou o som de uma cria, o input visual cai para zero (EAGLEMAN; VAUGHN, 2021). Dada a velocidade da plasticidade cortical — mensurável em menos de uma hora —, o córtex visual correria o risco de ser tomado por outras modalidades sensoriais todas as noites (EAGLEMAN; VAUGHN, 2021).

A DAT propõe que o sono REM (Rapid Eye Movement) e o conteúdo visual dos sonhos evoluíram como um mecanismo de defesa biológica para manter o córtex visual ativo e protegido contra a incursão de sentidos concorrentes (EAGLEMAN; VAUGHN, 2021). Durante o sono REM, ondas ponto-geniculo-occipitais (PGO) bombardeiam o córtex visual com descargas elétricas precisas, gerando uma simulação visual endógena que atua como um "sinal de manutenção". Os sonhos, portanto, não seriam apenas para a consolidação da memória, mas um "screensaver" protetor que preserva o território funcional da visão (EAGLEMAN; VAUGHN, 2021).

Evidências Experimentais e Comparativas

Experimentos realizados em 2025 reforçaram a validade da DAT ao demonstrar a rapidez da colonização cortical. Participantes videntes submetidos a vendagens por apenas 40 a 60 minutos começaram a mostrar atividade no córtex visual em resposta a sons e ao toque (PORTIER, 2026). Ao quinto dia de privação visual, o córtex occipital já estava sendo recrutado para tarefas complexas como a leitura em Braille, o que indica que a plasticidade permanece como força constante e ativa (PORTIER, 2026).

Tabela 2. Comparação da plasticidade cortical e da proporção de sono REM entre espécies

Parâmetro de Comparação	**Espécies com Alta Plasticidade (ex.: humano)**	**Espécies com Baixa Plasticidade (ex.: macaco-vervet)**
Status ao Nascer	"Cérebro mal-passado"; altamente dependente do aprendizado ambiental (PORTIER, 2026)	"Pré-programado"; funções sensoriais e motoras quase completas ao nascer (PORTIER, 2026)
Proporção de Sono REM	Elevada (~21-25% do tempo de sono) (EAGLEMAN; VAUGHN, 2021)	Reduzida (~6% do tempo de sono) (PORTIER, 2026)

Parâmetro de Comparação	**Espécies com Alta Plasticidade (ex.: humano)**	**Espécies com Baixa Plasticidade (ex.: macaco-vervet)**
Necessidade de Defesa	Alta; risco constante de colonização de áreas não utilizadas (EAGLEMAN; VAUGHN, 2021)	Baixa; circuitos neuronais mais fixos e menos vulneráveis (KINESIS MAGAZINE, 2024)
Correlação Evolutiva	O sono REM aumenta em passo firme com a recência evolutiva humana (EAGLEMAN; VAUGHN, 2021)	O sono REM é vestigial ou limitado a funções básicas (EAGLEMAN; VAUGHN, 2021)

A análise de 25 espécies de primatas em 2025 confirmou que a quantidade de sono REM está diretamente correlacionada ao nível de plasticidade neuronal de cada espécie (EAGLEMAN; VAUGHN, 2021). Além disso, a teoria explica por que bebês humanos passam quase 50% do sono em estágio REM: seus sistemas visuais altamente plásticos exigem uma defesa considerável para estabilizar os circuitos antes que outros sentidos os dominem (PORTIER, 2026). No envelhecimento, à medida que a plasticidade diminui, a necessidade de defesa visual também decai, o que resulta na redução natural da proporção de sono REM observada em idosos (EAGLEMAN; VAUGHN, 2021). O *Locus Coeruleus* (LC) foi identificado em estudos de 2026 como o hub mecanicista que integra esses processos, mantendo a estabilidade do sono REM e a plasticidade cortical de forma consistente ao longo da vida adulta saudável (KNOPPER; HANSEN, 2023).

Hipótese do Cérebro Sobreajustado: Sonhos como Dados Aumentados

Embora a DAT explique *por que* o cérebro precisa de ativação noturna, a Hipótese do Cérebro Sobreajustado (Overfitted Brain Hypothesis - OBH), desenvolvida por Erik Hoel, explica *por que* essa ativação deve ser bizarra e surreal (PRINCE; RICHARDS, 2021). Em modelos de inteligência artificial, o sobreajuste (overfitting) ocorre quando um algoritmo aprende tão bem os detalhes específicos e o ruído do seu conjunto de treinamento que perde a capacidade de generalizar para novos dados (PRINCE; RICHARDS, 2021). Hoel propõe que o cérebro humano enfrenta o mesmo perigo devido à natureza repetitiva da vida cotidiana (PRINCE; RICHARDS, 2021).

A Função da Estranheza Onírica

O cotidiano repetitivo — dirigir pelo mesmo caminho, realizar as mesmas tarefas — cria um viés estatístico nos modelos mentais. Se o cérebro se ajustasse perfeitamente a essa rotina, ele se tornaria rígido e incapaz de lidar com situações imprevistas (PRINCE; RICHARDS, 2021). Os sonhos, ao gerarem estimulação sensorial "fora da distribuição" (out-of-distribution) e deliberadamente corrompida, agem como um mecanismo de regularização (PRINCE; RICHARDS, 2021). A estranheza dos sonhos não é um erro biológico, mas uma característica funcional: é a injeção de ruído estocástico que impede o colapso do sistema em modelos simplistas da realidade (PRINCE; RICHARDS, 2021).

Este processo é comparado ao "recozimento simulado" (simulated annealing) na computação, em que a introdução de aleatoriedade permite que o sistema escape de mínimos locais, isto é, soluções subótimas de pensamento, para encontrar um estado de generalização global (PRINCE; RICHARDS, 2021). Evidências de 2025 sugerem que o gatilho mais potente para sonhos visualmente intensos é justamente a prática recorrente de uma tarefa nova e repetitiva durante o dia, como o efeito Tetris (PRINCE; RICHARDS, 2021). O cérebro detecta o início do overfitting e ativa o "Gerador" de sonhos para desestabilizar o modelo e resgatar a generalizabilidade (PRINCE; RICHARDS, 2021).

Arte e Ficção como Sonhos Artificiais

Uma das contribuições mais notáveis da OBH nos últimos 18 meses é a reinterpretação da utilidade cognitiva das artes (PRINCE; RICHARDS, 2021). Hoel sugere que filmes, romances e obras de arte surrealistas funcionam como "substitutos de sonhos" (PRINCE; RICHARDS, 2021). Ao expor a mente a cenários fantásticos e narrativas que desafiam a lógica linear, a ficção cumpre a mesma função biológica de prevenir o sobreajuste e aumentar a resiliência cognitiva (PRINCE; RICHARDS, 2021). Em sociedades em que o sono é cronicamente privado, o consumo de ficção pode atuar como mecanismo compensatório importante para manter a flexibilidade mental (PRINCE; RICHARDS, 2021).

Tabela 3. Conceitos da Hipótese do Cérebro Sobreajustado e suas implicações para a saúde mental

Conceito OBH	**Definição Neurobiológica**	**Implicação para a Saúde Mental**

Conceito OBH	Definição Neurobiológica	Implicação para a Saúde Mental
Overfitting Cognitivo	Fixação em rotinas e padrões repetitivos que reduzem a capacidade de inovação (PRINCE; RICHARDS, 2021)	Aumento da ansiedade frente ao imprevisto e perda de criatividade (PRINCE; RICHARDS, 2021)
Dados Corrompidos	Conteúdo onírico bizarro que desafia as categorias mentais habituais (PRINCE; RICHARDS, 2021)	Fortalecimento da robustez mental e da capacidade de generalização (PRINCE; RICHARDS, 2021)
Ficção como Prótese	Uso de artes e histórias para induzir estados de "sonho acordado" (PRINCE; RICHARDS, 2021)	Ferramenta educacional para aumentar a empatia e a flexibilidade social (PRINCE; RICHARDS, 2021)

A Neurobiologia do Propósito
e a Resiliência Cognitiva

A busca por propósito emergiu em 2026 não apenas como conceito filosófico, mas como construto biológico mensurável sob o nome de "teleonome" (THE TELEONOME..., 2026). O teleonome é definido como o sistema integrado de capacidades perceptivas, fisiológicas e afetivas que orienta o organismo em direção a objetivos adaptativos (THE TELEONOME..., 2026). Em animais sencientes, este sistema permite que as experiências afetivas atuem como sinais reguladores para priorizar demandas e governar o aprendizado em ambientes dinâmicos (THE TELEONOME..., 2026).

O Mix de Contextos
e o Refinamento da Estratégia de Vida

A lógica adversarial estende-se à construção do propósito por meio do fenômeno de mistura de contextos oníricos. Quando o cérebro insere colegas de trabalho atuais em cenários de infância durante o sonho, ele não está apenas consolidando memórias; ele está testando a resiliência dos esquemas mentais do indivíduo. Se uma intenção ou valor, por exemplo, "ser um líder ético", sobrevive e permanece coerente mesmo em cenários oníricos adversos e caóticos, esse valor se fortalece como heurística norteadora para a vida desperta.

Esta "negociação adversarial" ininterrupta permite que o sentido da vida seja a solução ótima para uma equação de adaptação contínua (COGNITIVE NEUROSCIENCE SOCIETY, 2026). Estudos apresentados na *Cognitive Neuroscience Society* (CNS) em 2026, em Vancouver, revelaram que indivíduos com maior clareza de propósito apresentam uma reserva cognitiva significativamente maior (COGNITIVE NEUROSCIENCE SOCIETY, 2026). O propósito atua como o "Discriminador" principal, conferindo ordem ao caos perceptual e protegendo o cérebro contra o declínio funcional associado à idade.

Resiliência Cognitiva
no Envelhecimento

Pesquisas longitudinais de 2025 e 2026, como o estudo da coorte *Resilient Minds* (ReMind) com 1.640 adultos, identificaram fatores sociais e de estilo de vida que conferem imunidade neurocognitiva (ANSTEY et al., 2026). A religiosidade, o apoio social emocional e o sono de qualidade foram destacados como moderadores críticos que permitem ao cérebro idoso funcionar de forma comparável ao de jovens em ambientes naturais (COGNITIVE NEUROSCIENCE SOCIETY, 2026).

Tabela 4. Fatores de resiliência cognitiva e seus mecanismos neurobiológicos

Fator de Resiliência	Mecanismo Neurobiológico	Impacto no Envelhecimento
Senso de Propósito	Redução da vulnerabilidade a placas amiloides; manutenção da integridade da substância branca (AMERICAN PSYCHOLOGICAL ASSOCIATION, 2026)	Redução de 28% no risco de desenvolver demência em 15 anos (UC DAVIS HEALTH, 2025)
Atividade Física	Remodelação da conectividade ACC-Amígdala; fortalecimento da plasticidade	Neutralização dos efeitos neurobiológicos de traumas de infância

Fator de Resiliência	Mecanismo Neurobiológico	Impacto no Envelhecimento
	sináptica (EUREKALERT!, 2026)	(EUREKALERT!, 2026)
Apoio Social/Religioso	Modulação da carga alostática e do cortisol por meio de redes de segurança emocional (COGNITIVE NEUROSCIENCE SOCIETY, 2026)	Preservação das funções executivas e da memória de interferência (COGNITIVE NEUROSCIENCE SOCIETY, 2026)
Uso de Próteses (Aparelhos Auditivos)	Prevenção do vácuo sensorial que levaria ao sequestro cortical de áreas auditivas (COGNITIVE NEUROSCIENCE SOCIETY, 2026)	Manutenção da integridade do córtex temporal e prevenção da depressão (COGNITIVE NEUROSCIENCE SOCIETY, 2026)

A pesquisa de Karen Campbell revelou uma descoberta importante em 2026: enquanto tarefas laboratoriais artificiais mostram declínio acentuado em idosos, em condições naturais, como assistir a filmes, os cérebros de adultos jovens e idosos processam a informação de forma quase idêntica (COGNITIVE NEUROSCIENCE SOCIETY, 2026). Isso sugere que o cérebro idoso compensa a perda de velocidade de processamento com o uso de conhecimentos acumulados e estratégias semânticas preditivas, num exemplo de "estabilidade por meio da mudança" ou alostase regulada pelo modelo SENECA (GUICHET et al., 2026).

Avanços em IA Adversarial para Diagnóstico e Intervenção Clínica

O uso de arquiteturas adversárias transcendeu a teoria para produzir ferramentas diagnósticas e terapêuticas relevantes nos últimos 18 meses. Um dos maiores avanços ocorreu no estudo dos Distúrbios da Consciência (DOC) (PUBMED, 2026).

Framework Adversarial para Consciência e Coma

1. Em 2026, um estudo liderado por Daniel Toker introduziu um framework de IA adversarial que colocou redes neurais profundas, treinadas em mais de 680.000 amostras eletrofisiológicas, contra modelos cerebrais interpretáveis (PUBMED, 2026). Sem programação explícita, a IA conseguiu reverter a engenharia do coma, identificando dois mecanismos biológicos fundamentais que falham em pacientes em estado vegetativo:

2. **Disrupção da via indireta dos gânglios da base**, impedindo a modulação do tráfego neural entre o tálamo e o córtex (PUBMED, 2026).

Aumento patológico do acoplamento inibitório-para-inibitório, validado por sequenciamento de RNA, que "tranca" o cérebro em um estado de baixa complexidade (PUBMED, 2026).

Essas descobertas permitiram a identificação de novos alvos para a Estimulação Cerebral Profunda (DBS). Em termos translacionais, a estimulação de alta frequência no núcleo subtalâmico surge como candidata plausível para restaurar a dinâmica de "grid elétrico" necessária para a consciência, com resultados iniciais que sugerem recuperação parcial de seguimento de instruções e funções motoras básicas em pacientes anteriormente não responsivos (PUBMED, 2026).

Interface Cérebro-Conteúdo e Decodificação Generativa

A área de Interfaces Cérebro-Computador (BCI) evoluiu do reconhecimento de comandos simples para a reconstrução de conteúdos semânticos complexos por meio de GANs e Transformers (GUO et al., 2026). Em 2025, o uso de redes AC-GAN (Auxiliary Conditional GANs) permitiu a reconstrução de imagens visuais diretamente da atividade cerebral com 92% de precisão em testes de classificação de dígitos (TANNER; CUTSURIDIS, 2025). Além disso, a aplicação de modelos de difusão e redes neurais spiking (SNNs) foi descrita como alternativa promissora para a conversão de sinais neurais em fala humana, com melhor eficiência computacional (TANNER; CUTSURIDIS, 2025).

A síntese de imagens médicas também se beneficiou dessas arquiteturas. O uso de CycleGANs para transformar ressonâncias magnéticas de T1 para T2 permitiu a criação de "gêmeos cerebrais virtuais", facilitando o diagnóstico de Alzheimer e tumores cerebrais por meio de controles personalizados (GAN-BASED CROSS-MODALITY..., [s.d.]). Esta abordagem de "neurodiversidade de precisão" permite identificar subgrupos neurobiológicos em condições como autismo e TDAH que seriam invisíveis aos critérios diagnósticos convencionais (MOHAMMAD et al., 2025).

Guerra Cognitiva e Neural Dominion: A Ameaça da Plasticidade

A mesma característica que torna o cérebro humano adaptável — sua plasticidade sináptica elevada —

também o torna vulnerável a novas formas de agressão tecnológica (ĆOSIĆ et al., 2026), conforme a Figura 2. A literatura recente sobre guerra cognitiva descreve uma migração dos sistemas de governança e controle do ambiente externo para o próprio campo de percepção humano (ĆOSIĆ et al., 2026).

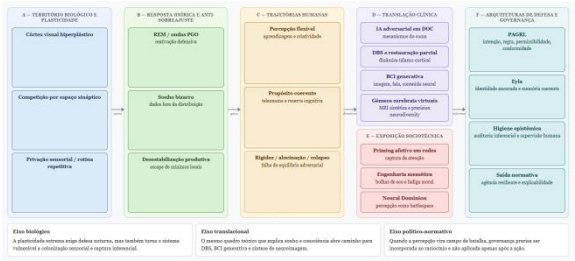

Figura 2. Vias de ataque da guerra cognitiva sobre o teleonome: captura das vias dopaminérgicas, indução de sobreajuste social (bolhas de eco) e resposta de governança pelos loops Eyla e PAGRL.

A Colonização da Percepção por Estímulos Digitais

Em 2025 e 2026, a literatura estratégica sobre guerra cognitiva alertou para o uso de mecanismos de priming, comunicação emocional e arquitetura algorítmica capazes de modular a emoção coletiva em tempo real (ĆOSIĆ et al., 2026). Ao sequestrar as vias de recompensa dopaminérgicas e injetar estímulos que mimetizam os sinais de sobrevivência do teleonome, atores externos podem fragmentar o senso de propósito de uma sociedade e induzir fadiga moral por meio de campanhas de desinformação (ĆOSIĆ et al., 2026). A mente humana tornou-se o novo "campo de batalha" (*battlespace*), em que a verdade e a ficção são borradas deliberadamente para minar a soberania cognitiva (ĆOSIĆ et al., 2026).

Adversários utilizam engenharia memética e IA persuasiva para criar "bolhas de eco" que funcionam como sobreajustes sociais artificiais, impedindo que os cidadãos generalizem informações fora de suas distribuições ideológicas (ĆOSIĆ et al., 2026). Este fenômeno é o equivalente social do cérebro sobreajustado de Hoel: a perda da flexibilidade cognitiva coletiva em favor de modelos rígidos e manipuláveis da realidade (PRINCE; RICHARDS, 2021; ĆOSIĆ et al., 2026).

Novas Arquiteturas de Defesa e Governança

1. Para combater essas ameaças, a neurociência teórica propôs, em 2026, novas arquiteturas de integridade. A arquitetura **Eyla** foca na consistência da identidade, integrando "priors biológicos" e módulos de memória HiPPO para permitir que agentes, humanos ou sintéticos, mantenham um self coerente mesmo sob pressão adversarial e *social engineering* (ARIF, 2026).

2. Complementarmente, o framework **PAGRL** (Pre-Action Governance Reasoning Loop) formaliza o processo de autogovernança comportamental. Inspirado nas funções executivas do córtex pré-frontal, o PAGRL exige que qualquer ação consequente passe por um loop de raciocínio em quatro estágios:

3. **Formação da Intenção**: identificação do impulso de agir (BANDARA et al., 2026).

4. **Recuperação de Regras**: acesso a um conjunto estruturado de normas éticas e situacionais (BANDARA et al., 2026).

5. **Raciocínio de Permissibilidade**: avaliação explícita da ação contra as regras em linguagem natural (BANDARA et al., 2026).

Decisão de Conformidade: execução, autocorreção ou escalonamento para supervisão humana (BANDARA et al., 2026).

Este modelo sugere que embutir a governança diretamente no processo de raciocínio — em vez de aplicá-la como filtro externo *post hoc* — produz comportamentos mais robustos e explicáveis frente a ambientes "fora da distribuição" e ataques adversários (BANDARA et al., 2026).

Limitações do Paradigma Adversarial e Auditabilidade

A utilidade do paradigma adversarial depende de uma fronteira conceitual clara: ele é uma matriz de leitura, não uma descrição literal de módulos anatômicos isolados (SHAW et al., 2026). A evidência consolidada sustenta que o cérebro opera por predição, erro, plasticidade e regulação dinâmica, mas a tradução desses processos para pares como Gerador e Discriminador permanece uma extrapolação heurística (SHAW et al., 2026; BUZEA; AGOP, 2025). Em termos metodológicos, a analogia com GANs ajuda a organizar relações entre simulação interna e teste de plausibilidade, porém não autoriza concluir que existam dois agentes cerebrais discretos, com objetivos formalmente idênticos aos de uma rede neural artificial. O valor do modelo está em explicitar tensões funcionais, não em substituir a complexidade distribuída do cérebro por uma arquitetura binária.

Essa ressalva é especialmente importante na discussão sobre lateralização hemisférica (BUZEA; AGOP, 2025). A associação do hemisfério direito a síntese narrativa e do hemisfério esquerdo a verificação analítica pode ser útil como inferência plausível, desde que apresentada como tendência funcional e não como dicotomia rígida (BUZEA; AGOP, 2025). Linguagem excessivamente simétrica corre o risco de apagar circuitos subcorticais, redes frontoparietais, sistemas afetivos e mecanismos de

neuromodulação que também participam da percepção, da imaginação e do propósito (SHAW et al., 2026; KNOPPER; HANSEN, 2023). Portanto, a arquitetura adversarial deve ser lida como camada interpretativa sobre processos distribuídos. Ela ganha rigor quando identifica quais achados são diretamente observados, quais são conectados por inferência e quais funcionam apenas como hipótese organizadora para pesquisa futura (SHAW et al., 2026).

A DAT e a OBH ilustram bem essa necessidade de calibração. A evidência consolidada para a DAT reside na relação entre plasticidade cortical, privação sensorial e ativação visual durante o sono REM; já a leitura dos sonhos como defesa territorial do córtex visual continua sendo uma hipótese mecanística que organiza esses achados (EAGLEMAN; VAUGHN, 2021; KNOPPER; HANSEN, 2023). De modo análogo, a OBH oferece uma inferência plausível ao aproximar sonhos, ruído sensorial e regularização contra sobreajuste, mas a equivalência entre sonho biológico e aumento artificial de dados deve permanecer explicitamente delimitada (PRINCE; RICHARDS, 2021). A contribuição do relatório, portanto, não é declarar que uma teoria elimina a outra, mas mostrar que ambas convergem na ideia de que a mente precisa produzir variação interna para preservar flexibilidade adaptativa (EAGLEMAN; VAUGHN, 2021; PRINCE; RICHARDS, 2021).

Também há limitações na passagem do nível individual ao nível social. A guerra cognitiva e a governança de agentes autônomos compartilham vocabulário com a neurociência porque lidam com atenção, crença, intenção e erro, mas seus mecanismos causais não são redutíveis aos mesmos processos sinápticos (ĆOSIĆ et al., 2026; BANDARA et al., 2026). Quando se afirma que bolhas de eco funcionam como sobreajustes sociais, a formulação deve ser entendida como analogia operacional: comunidades informacionais podem perder capacidade de generalização, mas não "sonham" nem "regularizam" seus modelos da mesma forma que organismos biológicos. Essa distinção protege o texto contra reducionismo e contra a tentação de converter uma metáfora forte em diagnóstico totalizante (ĆOSIĆ et al., 2026).

No eixo clínico, a prudência é ainda mais necessária (PUBMED, 2026; GUO et al., 2026). A aplicação de IA adversarial em distúrbios da consciência, reconstrução de imagens cerebrais e síntese de exames médicos aponta um campo translacional relevante, mas a passagem de desempenho algorítmico para intervenção terapêutica exige validação independente, amostras diversas e explicabilidade suficiente para orientar decisão médica (PUBMED, 2026; GUO et al., 2026; TANNER; CUTSURIDIS, 2025; MOHAMMAD et al., 2025). Métricas altas de classificação ou reconstrução não equivalem automaticamente a compreensão causal (GUO et al., 2026; MOHAMMAD et al., 2025). O relatório, assim, trata esses avanços como fronteiras promissoras, não como evidência conclusiva de que arquiteturas adversárias já resolvem a clínica neuropsiquiátrica.

Para preservar auditabilidade, cada afirmação central deve ser lida em três níveis. Primeiro, a fonte primária: artigo revisado por pares, preprint, material institucional ou documento interno. Segundo, o tipo de operação lógica: descrição empírica, inferência plausível, extrapolação heurística ou recomendação normativa. Terceiro, o grau de consequência: se a afirmação apenas reorganiza a interpretação do fenômeno, se sugere um programa de pesquisa ou se sustenta implicações clínicas e políticas. Essa matriz evita que fontes de divulgação recebam o mesmo peso que estudos revisados por pares e torna visível quando o texto migra de mecanismo neurobiológico para governança social (EAGLEMAN; VAUGHN, 2021; SHAW et al., 2026; BANDARA et al., 2026).

Por fim, a ferramenta local validator_logico.py integra esse ciclo de auditoria como apoio operacional, não como substituto da revisão humana. Ela identifica densidade citacional, duplicidades bibliográficas, presença de seção de limitações, equilíbrio entre termos assertivos e cautelosos, e dependência de fonte interna. Esses sinais não decidem a verdade das teses, mas ajudam a localizar onde o relatório precisa de qualificação, citação ou separação entre fato e interpretação. A robustez final depende justamente da tensão adversarial que o próprio texto descreve: a geração de hipóteses ousadas deve ser acompanhada por discriminação crítica, controle de evidência e disposição para revisar a arquitetura conceitual quando a literatura exigir (BANDARA et al., 2026).

CONCLUSÃO

O balanço dos últimos dezoito meses sugere que o cérebro humano não busca apenas repouso, mas também persistência competitiva. A percepção visual, a estranheza dos sonhos e a busca por propósito compartilham, neste enquadramento, uma mesma raiz funcional: a necessidade de preservar território neural e flexibilidade adaptativa.

Enquanto a Teoria da Ativação Defensiva nos mostra o cérebro protegendo seu hardware sensorial contra a colonização noturna, a Hipótese do Cérebro Sobreajustado nos mostra o sistema protegendo seu software cognitivo contra a rigidez da rotina. O propósito, nesse contexto, emerge como a bússola que organiza essas tensões adversárias, permitindo que a identidade humana sobreviva ao caos onírico e à manipulação digital externa.

A saúde mental e a soberania cognitiva no século XXI dependem da integridade desses mecanismos de defesa e geração (PUBMED, 2026; COGNITIVE NEUROSCIENCE SOCIETY, 2026; ĆOSIĆ et al., 2026). Os avanços em IA adversarial clínica e neurotecnologia de 2026 abrem um campo terapêutico promissor para distúrbios da consciência e doenças neurodegenerativas, mas também exigem nova ética da percepção (PUBMED, 2026; GUO et al., 2026; BANDARA et al., 2026). Neste quadro teórico, o "sentido da vida" pode ser descrito como solução adaptativa provisória para uma equação que nunca cessa (THE TELEONOME..., 2026). O futuro da neurocognição reside menos em eliminar a fricção entre sonho, erro e propósito do que em compreender como essa fricção pode ser convertida em crescimento, discernimento e resiliência.

AGRADECIMENTOS

Os autores agradecem à Secretaria de Ciência, Tecnologia e Inovação em Saúde (SCTIE) e aos colegas revisores cujas contribuições aprimoraram a calibração metodológica deste trabalho.

REFERÊNCIAS

AMERICAN PSYCHOLOGICAL ASSOCIATION. *Sense of purpose in life may preserve cognitive function, and more scientific findings*. 2026. Disponível em: https://www.apa.org/monitor/2026/01-02/research-summaries-purpose-cognition. Acesso em: 14 maio 2026.

ANSTEY, Kaarin J. et al. Cohort profile: the Resilient Minds national study of mental health and cognitive resilience in community dwelling adults aged 18 to 93. *Frontiers in Digital Health*, v. 8, art. 1710349, 2026. DOI: 10.3389/fdgth.2026.1710349. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC12989522/. Acesso em: 14 maio 2026.

ARIF, Aditto. *Eyla: toward an identity-anchored LLM architecture with integrated biological priors vision, implementation attempt, and lessons from AI-assisted development*. arXiv, 2026. arXiv:2604.00009v1. Disponível em: https://arxiv.org/html/2604.00009v1. Acesso em: 14 maio 2026.

BANDARA, Eranga et al. *Think before you act: a neurocognitive governance model for autonomous AI agents*. arXiv, 2026. arXiv:2604.25684v1. Disponível em: https://arxiv.org/html/2604.25684v1. Acesso em: 14 maio 2026.

BUZEA, Calin Gheorge; AGOP, Maricel. *The creative brain: a generative adversarial network-inspired exploration of mind and reality*. Hauppauge, NY: Nova Science Publishers, 2025. DOI: 10.52305/BTLG0887. Disponível em: https://novapublishers.com/shop/the-creative-brain-a-generative-adversarial-network-inspired-exploration-of-mind-and-reality/. Acesso em: 14 maio 2026.

COGNITIVE NEUROSCIENCE SOCIETY. *Modeling brain aging and resilience over the lifespan reveals new individual factors*. 2026. Disponível em: https://www.cogneurosociety.org/modeling-brain-aging-and-resilience-over-the-lifespan-reveals-new-individual-factors/. Acesso em: 14 maio 2026.

ĆOSIĆ, Krešimir; POPOVIĆ, Siniša; WIEDERHOLD, Brenda. Emotionally based strategic communications as a new tool in defensive cognitive warfare. *Frontiers in Psychology*, v. 17, art. 1751406, 2026. DOI: 10.3389/fpsyg.2026.1751406. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC12920201/. Acesso em: 14 maio 2026.

EAGLEMAN, David M.; VAUGHN, Don A. The defensive activation theory: REM sleep as a mechanism to prevent takeover of the visual cortex. *Frontiers in Neuroscience*, v. 15, art. 632853, 2021. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC8176926/. Acesso em: 14 maio 2026.

EUREKALERT! *Novel study shows physical activity reshapes neural connectivity and makes the brain more resilient after childhood trauma*. Washington, DC, 17 mar. 2026. DOI: 10.1016/j.bpsc.2026.01.006. Disponível em: https://www.eurekalert.org/news-releases/1120255. Acesso em: 14 maio 2026.

GAN-BASED CROSS-MODALITY BRAIN MRI SYNTHESIS: paired versus unpaired training and comparison with diffusion and transformer models. *Biomimetics*, v. 11, n. 3, art. 175, [s. d.]. Disponível em: https://www.mdpi.com/2313-7673/11/3/175. Acesso em: 14 maio 2026.

GUICHET, Clément et al. Cognitive flexibility and brain network energy in healthy aging: an allostatic perspective from the SENECA model. *Imaging Neuroscience*, v. 4, art. IMAG.a.1091, 2026. DOI: 10.1162/IMAG.a.1091. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC12797148/. Acesso em: 14 maio 2026.

GUO, Yi et al. Generative AI for brain-computer interfaces decoding: advances, challenges and future. *The Innovation Medicine*, v. 4, art. 100193, 2026. DOI: 10.59717/j.xinn-med.2026.100193. Disponível em: https://www.the-innovation.org/article/doi/10.59717/j.xinn-med.2026.100193. Acesso em: 14 maio 2026.

KINESIS MAGAZINE. *Why do we dream?: exploring a new theory*. 2024. Disponível em:

https://kinesismagazine.com/2024/01/12/why-do-we-dream-exploring-a-new-theory/. Acesso em: 14 maio 2026.

KNOPPER, Rasmus West; HANSEN, Brian. Locus coeruleus and the defensive activation theory of rapid eye movement sleep: a mechanistic perspective. *Frontiers in Neuroscience*, v. 17, art. 1094812, 2023. DOI: 10.3389/fnins.2023.1094812. Disponível em: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1094812/full. Acesso em: 14 maio 2026.

MOHAMMAD, Suleiman Ibrahim et al. Precision neurodiversity: personalized brain network architecture as a window into cognitive variability. *Frontiers in Human Neuroscience*, v. 19, art. 1669431, 2025. DOI: 10.3389/fnhum.2025.1669431. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC12647089/. Acesso em: 14 maio 2026.

PORTIER, Lauren. New theory suggests dreams are the brain's defense against going blind. *Cleveland 13 News*, Cleveland, 23 abr. 2026. Disponível em: https://www.cleveland13news.com/story/new-theory-suggests-dreams-are-the-brain-s-defense-against-going-blind. Acesso em: 14 maio 2026.

PRINCE, Luke Y.; RICHARDS, Blake A. The overfitted brain hypothesis. *Patterns*, v. 2, n. 5, art. 100268, 2021. DOI: 10.1016/j.patter.2021.100268. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC8134936/. Acesso em: 14 maio 2026.

PUBMED. *Adversarial AI reveals mechanisms and treatments for disorders of consciousness*. 2026. Disponível em: https://pubmed.ncbi.nlm.nih.gov/41876853/. Acesso em: 14 maio 2026.

SHAW, Alexander D.; SUMNER, Rachael L.; BERNDT, Lioba C. S. Predictive coding and neurocomputational psychiatry: a mechanistic framework for understanding mental disorders. *Frontiers in Psychiatry*, v. 16, art. 1713833, 2026. DOI: 10.3389/fpsyt.2025.1713833. Disponível em: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2025.1713833/full. Acesso em: 14 maio 2026.

TANNER, Tim; CUTSURIDIS, Vassilis. Generative adversarial network for image reconstruction from human brain activity. In: INTERNATIONAL JOINT CONFERENCE ON BIOMEDICAL ENGINEERING SYSTEMS AND TECHNOLOGIES, 18., 2025. *Proceedings* [...]. [S. l.]: SciTePress, 2025. p. 868-877. DOI: 10.5220/0013149300003911. Disponível em: https://www.scitepress.org/PublishedPapers/2025/131493/. Acesso em: 14 maio 2026.

THE TELEONOME: a framework for understanding animal welfare integrating adaptive capabilities, affective regulation, agency, and environmental affordances. *Frontiers*, 2026. Disponível em: https://www.frontiersin.org/journals/animal-science/articles/10.3389/fanim.2026.1768519/full. Acesso em: 14 maio 2026.

UC DAVIS HEALTH. *Having a sense of purpose may protect against dementia*. 2025. Disponível em: https://health.ucdavis.edu/news/headlines/having-a-sense-of-purpose-may-protect-against-dementia/2025/08. Acesso em: 14 maio 2026.

VARUGHESE, Jobit. What are generative adversarial networks (GANs)? *IBM Think*, [S. l.], [s. d.]. Disponível em: https://www.ibm.com/think/topics/generative-adversarial-networks. Acesso em: 14 maio 2026.

ZHAO, Tingting et al. *Generative adversarial networks (GAN): from theory to practice*. London: IntechOpen, 2026. Disponível em: https://www.intechopen.com/welcome/1005517. Acesso em: 14 maio 2026.