

Vitor Martins Dutra

# Inteligência ARTIFICIAL

A era dos multiagentes de  
IA generativa no Direito

2026

 EDITORA  
*Jus*PODIVM  
[www.editorajuspodivm.com.br](http://www.editorajuspodivm.com.br)

# 5

## MULTIAGENTES E ORQUESTRAÇÃO DA IA

### 5.1. A LINHA EVOLUTIVA COMPLETA: DE TAREFAS A SISTEMAS

Chegamos, enfim, ao capítulo tecnicamente mais instigante.

E como você já deve ter percebido, desde a primeira página da presente obra, busquei lhe conduzir por um aprendizado sucessivo e progressivo, proporcionando camadas de conteúdo que se conectam, em especial no presente capítulo. Espero que, agora, tudo faça sentido (se ainda não o fez).

A essa altura talvez já esteja cansado de repetir alguns conceitos, mas acredite, eles são necessários para consolidar o entendimento que, agora, ficará ainda mais claro. Em suma, este é o capítulo que os capítulos anteriores prepararam.

Cada conceito apresentado até aqui foi um degrau na escada que nos traz a este ponto. E este ponto é, sem exagero, o lugar onde o futuro da prática jurídica está sendo definido.

O lugar a que me referi no parágrafo anterior são os sistemas multiagênticos: a orquestração de inteligências artificiais que trabalham em conjunto para resolver problemas que nenhum agente isolado conseguiria resolver com a mesma eficiência, a mesma consistência e a mesma escala.

Ainda vai demorar algum tempo para os profissionais perceberem o que você está percebendo agora! Aproveite! Você começa em vantagem.

Para situar o leitor com precisão, vale reconstituir a linha evolutiva completa que o livro percorreu. Acompanhe com atenção.

- 1) IA por tarefas: o profissional formula uma pergunta ou instrução, o modelo responde, o profissional avalia e segue em frente. Cada interação é uma unidade fechada, sem conexão estrutural com a anterior ou a seguinte. Este estágio já é familiar para a maioria dos juristas que usam IA generativa e foi o tema central da minha obra anterior.
- 2) IA em fluxos de trabalho com orquestração humana (infraestrutura cognitiva): os prompts deixam de ser pontuais e passam a integrar sequências deliberadas. O profissional humano é o orquestrador e é ele quem define a sequência, decide quando acionar cada componente e consolida os resultados. A infraestrutura cognitiva, descrita no capítulo anterior, opera neste estágio. A inteligência está distribuída entre o humano e os agentes, mas o governo do processo é inteiramente humano.
- 3) IA em sistemas multiagênticos: a diferença em relação ao estágio anterior não é apenas a quantidade de agentes envolvidos. É a natureza da interação entre eles. Eis um grande salto.

Num sistema multiagêntico, os agentes não apenas executam tarefas em sequência; eles se comunicam, dividem trabalho, verificam o trabalho uns dos outros e se coordenam para convergir em um resultado que nenhum deles teria produzido sozinho.

A orquestração pode ser humana, parcial ou totalmente automatizada por uma IA. Cada uma dessas configurações tem implicações diferentes para o controle, a responsabilidade e o potencial de escala do sistema.

Há, portanto, dentro deste terceiro estágio, um contínuo que vai da orquestração puramente humana à orquestração puramente automatizada, com múltiplas configurações híbridas entre esses extremos. Esse contínuo não é apenas uma possibilidade técnica: é um espectro de escolhas profissionais e éticas que o advogado precisa fazer conscientemente.

Um sistema multiagêntico bem projetado não é simplesmente a soma das capacidades dos seus agentes individuais. A divisão de trabalho cognitivo, a especialização profunda, a checagem cruzada entre agentes e a execução paralela criam um sistema cujo desempenho coletivo excede o que qualquer configuração de agente único poderia alcançar.

Já é possível afirmar que sistemas multiagênticos superam agentes únicos em tarefas complexas que exigem múltiplas habilidades ou perspectivas (WANG et al., 2024).

Agora é hora de se despir dos conceitos estanques e das lógicas acabadas para refletir sobre o todo e sobre a grande inspiração que quero incutir em você: compreender que um sistema agêntico ou multiagêntico é uma arquitetura e não uma ferramenta específica. Que, acima de tudo, é uma forma de enxergar a aplicação da IA no Direito, permitindo que você crie verdadeiras estruturas cognitivas que simulam seus fluxos de trabalho.

O jurista que compreende os três estágios acima está equipado para fazer a escolha adequada em cada situação, em vez de buscar sempre a solução mais sofisticada ou, inversamente, ficar preso ao primeiro estágio por desconhecimento dos demais.

Agora a grande virada que eu espero de você é permitir-se compreender a lógica como um todo. E é sobre isso que vou me esmerar a partir daqui.

## A LINHA EVOLUTIVA COMPLETA DO USO DA IA NO DIREITO



Vitor Martins Dutra - Multiagentes de IA no Direito

**Figura 24:** linha evolutiva proposta no uso da IA no Direito.

## 5.2. O QUE EMERGE NOS SISTEMAS MULTIAGÊNTICOS?

Por que não usar simplesmente um agente muito bom, com um prompt de sistema muito bem elaborado, em vez de construir um sistema com múltiplos agentes? Parece muito mais prático e simples, não é verdade?

A pergunta é legítima e merece uma resposta precisa, porque a complexidade adicional de um sistema multiagêntico só se justifica quando os benefícios superam os custos.

A primeira razão é a especialização. Um modelo de linguagem, por mais capaz que seja, tende a produzir melhores resultados quando opera dentro de um escopo bem delimitado.

Um agente configurado para analisar laudos periciais conhece profundamente as nuances desse tipo de documento. Um agente configurado para redigir petições iniciais domina os requisitos formais e as estratégias argumentativas dessa peça. Quando esses dois agentes trabalham em sequência (ou paralelamente), o primeiro identificando as fragilidades do laudo e o segundo construindo a argumentação a partir dessas fragilidades, o resultado é mais preciso do que o de um agente único encarregado de fazer tudo.

A especialização funciona na IA pelo mesmo motivo que funciona nas equipes humanas: foco produz excelência em um escopo específico (HONG et al., 2023).

A segunda razão é a verificação cruzada.

Em um sistema com múltiplos agentes, é possível configurar agentes cuja função específica é revisar o trabalho de outros agentes (validação e conferência). Um agente de redação produz uma petição; um agente de revisão verifica se a petição atende aos requisitos formais, se a fundamentação é consistente com a jurisprudência citada e se a linguagem está adequada aos padrões do escritório.

Essa camada de verificação automatizada não substitui a revisão humana final, mas filtra erros que o agente de redação não seria capaz de identificar no próprio trabalho, pelo mesmo motivo que um advogado que revisa seu próprio texto nem sempre percebe as falhas que outra pessoa perceberia imediatamente.

A terceira razão é a execução paralela.

Tarefas que em um sistema de agente único precisariam ser executadas em sequência podem, em um sistema multiagêntico, ser distribuídas e executadas simultaneamente. Enquanto um agente pesquisa jurisprudência, outro analisa a documentação, um terceiro verifica os prazos processuais e um quarto consulta a memória institucional sobre casos similares. Quando todos terminam, um agente consolidador reúne os insumos e produz o resultado integrado.

O tempo total de processamento cai de forma significativa e a qualidade do resultado final tende a ser superior porque cada componente foi produzido por um agente operando com foco total em sua tarefa específica. Não se olvide, como veremos na sequência, que a própria tarefa de supervisão humana fica facilitada, pois o conteúdo entregue de forma segmentada pelos agentes organiza a informação.

A quarta razão é a emergência de raciocínio coletivo.

Pesquisadores da Universidade Stanford demonstraram, em um experimento amplamente citado, que agentes de IA interagindo em um ambiente simulado desenvolviam comportamentos sociais complexos que não estavam previstos nas instruções individuais de nenhum deles: formavam grupos, disseminavam informações e coordenavam ações coletivas (PARK et al., 2023).

Embora esse experimento fosse uma simulação social, o princípio subjacente é o mesmo para o nosso caso: quando agentes de IA interagem de forma estruturada, propriedades emergem no nível do sistema que não existem no nível dos componentes individuais.

Em suma, um sistema multiagêntico não é apenas mais rápido ou mais eficiente que um agente único, mas pode ser qualitativamente diferente em sua capacidade de produzir análises multidimensionais, identificar inconsistências, integrar perspectivas distintas e gerar produtos de trabalho com um nível de completude que a operação individual, humana ou artificial, dificilmente iguala. Ao ter mais contexto, pode ser ainda mais completo.

### 5.3. ANATOMIA DE UM SISTEMA MULTIAGÊNICO

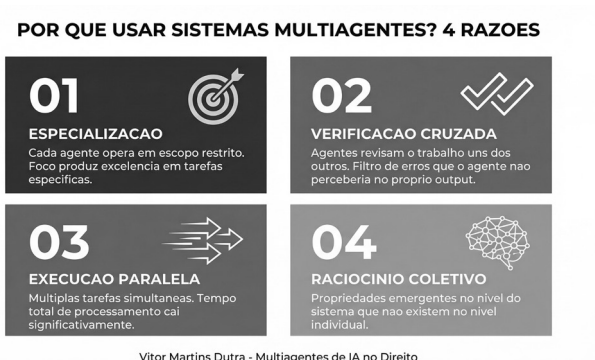
Um sistema multiagênico, independentemente de sua implementação tecnológica específica, é composto por um conjunto de elementos estruturais que se repetem em toda arquitetura funcional: agentes especialistas, um mecanismo de orquestração, protocolos de comunicação, memória compartilhada e pontos de intervenção humana. Vamos trabalhar esses conceitos.

#### 5.3.1. Agentes especialistas

Os agentes especialistas são as unidades operacionais do sistema. Cada um é configurado com um prompt de sistema que define sua especialização, seu escopo de atuação, as ferramentas a que tem acesso e as normas de comportamento que deve seguir.

No contexto jurídico, a especialização pode seguir diversas lógicas: por tipo de tarefa (pesquisa, redação, revisão, cálculo), por matéria jurídica (previdenciário, trabalhista, tributário), por etapa processual (petição inicial, contestação, recurso) ou por função dentro do escritório ou gabinete (atendimento, análise estratégica, produção de documentos).

A granularidade da especialização é uma decisão de projeto com consequências diretas sobre o desempenho e a manutenção do sistema. Agentes com escopo muito amplo perdem as vantagens da especialização. Agentes com escopo excessivamente estreito multiplicam a complexidade da orquestração sem ganho proporcional de qualidade. O equilíbrio depende do volume de trabalho do ambiente, da recorrência dos tipos de caso e da capacidade de manutenção dos agentes ao longo do tempo. É uma decisão de negócios.



**Figura 25:** razões para os sistemas multiagentes serem o grande potencial de uso da IA no Direito.

### 5.3.2. Mecanismo de orquestração e padrões de coordenação

O orquestrador é o componente que transforma um conjunto de agentes em um sistema coeso. Sua função é receber objetivos, decompô-los em subtarefas, atribuir cada subtarefa ao agente adequado, gerenciar a sequência e o paralelismo das operações, consolidar os resultados e determinar quando o objetivo foi alcançado ou quando é necessário escalar para intervenção humana.

O orquestrador pode ser um humano, um sistema de regras predefinidas ou um agente de IA cuja especialidade é precisamente a coordenação de outros agentes.

Cada abordagem tem perfis de risco, controle e eficiência diferentes, e trataremos da orquestração por IA em seção específica adiante neste capítulo.

Independentemente de quem ou o que orquestra, há quatro padrões de orquestração que se repetem na literatura e na prática: sequencial, paralelo, hierárquico e reflexivo.

No padrão sequencial, cada agente opera em ordem definida, passando o resultado ao próximo. No padrão paralelo, múltiplos agentes trabalham simultaneamente e os resultados são consolidados ao final. No padrão hierárquico, um agente gerente coordena agentes subordinados, podendo redistribuir tarefas conforme os resultados intermediários. No padrão reflexivo, o sistema inclui ciclos de autocorreção em que o output de um agente é submetido a outro para revisão antes de avançar (WU et al., 2023; HONG et al., 2023).

Na prática jurídica, os fluxos mais robustos podem combinar esses padrões.

Um caso de aposentadoria especial pode começar com análise paralela do CNIS e da documentação médica, seguida por uma fase sequencial de identificação de benefícios e redação da petição, com uma etapa reflexiva de revisão automatizada antes da entrega ao advogado. O orquestrador gerencia essa combinação e garante que cada transição entre fases ocorra no momento correto.



**Figura 26:** padrões de orquestração.

### 5.3.3. Protocolos de comunicação e “handoff”

Um dos desafios mais subestimados na construção de sistemas multiagênticos é a comunicação entre agentes. Li et al. (2023) demonstraram experimentalmente que a qualidade da comunicação entre agentes é um dos fatores mais determinantes do desempenho de sistemas multiagênticos, muitas vezes mais relevante que a capacidade individual de cada agente.

Quando o agente de pesquisa entrega seus resultados ao agente de redação, o que exatamente é entregue? O texto bruto? Um resumo? Uma lista estruturada?

A qualidade e o formato dessa comunicação determinam, em grande parte, a qualidade do trabalho do agente receptor.

Protocolos de comunicação bem definidos estabelecem o formato, o conteúdo e a estrutura da informação transmitida entre agentes. Em termos práticos, isso significa que cada agente não apenas produz seu resultado principal, mas o formata de acordo com as necessidades do agente que o receberá. Essa camada de padronização é o que permite que agentes desenvolvidos independentemente operem de forma coesa dentro do mesmo sistema.

O “handoff” é a transferência formal de trabalho de um agente para outro. É seguramente o momento mais crítico do fluxo. Se bem gerenciado preserva todo o contexto relevante, inclusive os metadados sobre como foi produzido, o que pode ser útil para disparar gatilhos de automação por exemplo.



**Figura 27:** exemplo prático de um orquestrador.

#### 5.3.4. Memória compartilhada

Em sistemas mais sofisticados, os agentes não se comunicam apenas por “handoffs” diretos: compartilham uma memória comum que funciona como repositório de contexto acessível a todos os componentes do sistema.

Essa memória compartilhada pode conter os dados do caso em processamento, os resultados parciais já produzidos por outros agentes, as decisões de roteamento do orquestrador e, em sistemas com memória de longo prazo, o histórico de casos anteriores e os padrões identificados pela infraestrutura cognitiva.

A vantagem da memória compartilhada é que ela elimina a necessidade de cada agente reconstruir o contexto a partir de informações fragmentadas. O agente de redação pode consultar diretamente os resultados da pesquisa jurisprudencial, os dados da análise do CNIS e as notas estratégicas produzidas pelo agente de triagem, tudo isso sem depender de que cada um desses agentes tenha formatado seu output especificamente para ele.

O framework AIOS endereça especificamente essa questão com seu gerenciador de memória e seu gerenciador de contexto (MEI et al., 2024).

### 5.3.5. Pontos de intervenção humana

Nenhum sistema multiagêntico voltado para a prática jurídica deve operar sem pontos definidos de intervenção humana.

Esses pontos são escolhas de design e devem ser posicionados nas etapas em que o julgamento profissional é insubstituível, em que o risco de erro tem consequências graves ou em que a comunicação com o usuário dos serviços exige a presença do advogado ou demais operadores.

A definição dos pontos de intervenção é uma decisão estratégica e normativa. Pontos demais eliminam os ganhos de eficiência do sistema. Pontos de menos expõem o escritório ou gabinete a riscos profissionais e éticos.

Um princípio útil é que o sistema deve ser projetado para operar com a menor intervenção humana possível, desde que toda intervenção necessária esteja garantida. A diferença entre essa formulação e a de “sem intervenção humana” não é semântica, mas é a diferença entre um sistema responsável e um sistema temerário (ou violador de normas).

### 5.4. O ORQUESTRADOR HUMANO: O JURISTA COMO MAESTRO

No modelo de orquestração humana, que é o ponto de partida mais seguro e recomendável para qualquer escritório ou gabinete que inicie sua jornada em sistemas multiagênticos, o operador responsável assume o papel de maestro do sistema.

Ele não executa as tarefas, mas coordena quem as executa, define a sequência, avalia os produtos intermediários e toma as decisões que exigem julgamento profissional. Essa posição exige competências específicas que transcendem o conhecimento jurídico tradicional.

O jurista-maestro precisa compreender as capacidades e as limitações de cada agente do sistema, saber identificar quando um output (resultado) intermediário tem qualidade suficiente para avançar ao próximo passo e quando precisa de ajuste, e ser capaz de reconfigurar o fluxo em tempo real quando um caso apresenta características que desviam do padrão esperado.

A analogia com o maestro de uma orquestra é ilustrativa. O maestro não toca nenhum instrumento durante o concerto, mas sem ele a orquestra não produz música coerente. Ele conhece a partitura, sabe quando um solista precisa ser contido e quando precisa ser projetado e toma decisões em tempo real que mantêm o conjunto coeso. O jurista que orchestra um sistema multiagêntico opera com a mesma lógica: não produz diretamente os resultados, mas garante que cada componente opera no momento certo, com o insumo certo e dentro dos parâmetros de qualidade adequados.

Uma distinção fundamental nesse papel é a diferença entre decidir e executar.

No modelo tradicional de trabalho jurídico, o advogado ou operador do Direito frequentemente acumula as duas funções: decide a estratégia e executa o trabalho de produção.

No modelo multiagêntico, a execução é delegada ao sistema, mantida a decisão com o profissional. Essa separação libera uma parcela significativa do tempo e da atenção do advogado para aquilo que mais precisa de sua inteligência: a análise estratégica do caso, a relação com o cliente, a avaliação de riscos e a tomada de decisões em situações que o sistema não está preparado para resolver autonomamente.

O perfil de competências do jurista-maestro inclui, pelo menos, três camadas.

A primeira é o conhecimento jurídico, que permanece essencial e insubstituível. A segunda é a compreensão dos sistemas: o que um agente pode e não pode fazer, como os fluxos operam, onde estão os pontos de fragilidade e como identificar quando o sistema está produzindo resultados abaixo do esperado. A terceira é a capacidade de projeto: saber conceber novos fluxos para novos tipos de caso, adaptar fluxos existentes quando as circunstâncias mudam e evoluir a infraestrutura cognitiva do escritório ao longo do tempo.

Novas habilidades que o profissional do Direito não está, em regra, habituado, mas que começa a se tornar essencial. E a combinação dessas três camadas define o que podemos chamar de jurista de nova geração: não um profissional que abandonou o Direito em favor da tecnologia, mas um profissional que usa a tecnologia para fazer Direito com uma escala, uma consistência e uma profundidade que o modelo anterior de trabalho não permitia.



**Figura 28:** camadas essenciais ao jurista maestro.

## 5.5. O ESTADO DA ARTE CIENTÍFICO

Wang et al. (2024), em uma das revisões mais abrangentes publicadas sobre sistemas multiagentes, mapearam a arquitetura geral dos agentes baseados em LLMs e identificaram quatro módulos fundamentais: perfil (quem o agente é), memória (o que o agente lembra), planejamento (como o agente decide o que fazer) e ação (como o agente executa). Os autores observaram que a integração desses quatro módulos é o que permite a emergência de comportamentos sofisticados, e que deficiências em qualquer um deles compromete o desempenho do sistema como um todo.

Park et al. (2023), já mencionados na presente obra, conduziram o experimento das “simulações generativas”, criando uma comunidade virtual com 25 agentes de IA dotados de memória, personalidade e objetivos individuais. Os agentes desenvolveram comportamentos sociais emergentes, incluindo disseminação de informações, formação de relações e coordenação de ações coletivas, nenhum dos quais estava explicitamente programado. O experimento demonstrou que agentes com memória e objetivos são capazes de interação social produtiva, com propriedades emergentes que não existiam no nível individual.

Wu et al. (2023), no artigo que introduziu o AutoGen, demonstraram que sistemas de conversa entre múltiplos agentes superaram consistentemente agentes individuais em tarefas que exigem raciocínio em múltiplos passos, verificação de resultados e integração de perspectivas diversas. Um achado relevante foi que a inclusão de feedback humano em pontos estratégicos do fluxo multiagêntico melhorava significativamente a qualidade dos resultados, sugerindo que o modelo ótimo não é nem totalmente automatizado nem totalmente humano, mas uma combinação calibrada de ambos.

Qian et al. (2023), no projeto ChatDev, demonstraram que agentes organizados em uma estrutura empresarial simulada, com papéis como CEO, CTO, programador e testador, eram capazes de produzir software funcional de forma colaborativa, completando ciclos de desenvolvimento completos. Embora o contexto seja de software, o princípio é diretamente transferível: quando agentes especializados colaboram dentro de uma estrutura organizacional bem definida, o produto coletivo é substancialmente superior ao que qualquer agente individual produziria.

Li et al. (2023), no projeto CAMEL, investigaram como agentes de IA podem colaborar para completar tarefas através de conversas estruturadas com atribuição clara de papéis. O estudo introduziu o conceito de “sociedade de mentes” (society of mind) aplicado a LLMs, em que agentes diferentes contribuem com perspectivas complementares que, integradas, produzem soluções mais robustas. Os experimentos mostraram que a comunicação estruturada entre agentes, com papéis bem definidos e protocolos de handoff claros, é significativamente mais eficaz do que a comunicação livre ou desestruturada.

Shinn et al. (2023) propuseram o framework Reflexion, em que um agente avalia verbalmente seus próprios erros após cada tentativa e usa essa avaliação para guiar a próxima iteração. Os resultados empíricos demonstraram melhoras substanciais de desempenho em tarefas de raciocínio sequencial, codificação e tomada de decisão, com o agente convergindo progressivamente para resultados corretos ao longo de múltiplas iterações.

Du et al. (2023) demonstraram que sistemas em que múltiplos agentes debatem uma questão, com cada agente argumentando sua posição e refinando-a em resposta às críticas dos outros, produzem respostas de qualidade significativamente superior à de um agente único, especialmente em tarefas que envolvem raciocínio factual e argumentativo.

Wang et al. (2023), no projeto Voyager, demonstraram que um agente equipado com memória de habilidades adquiridas e um mecanismo de síntese de novas habilidades a partir da experiência podia aprender progressivamente a resolver tarefas mais complexas sem supervisão humana durante o processo de aprendizado.

Sobre os riscos, é preciso documentar os seguintes achados.

Uma linha crescente de trabalhos documenta os riscos específicos que emergem quando múltiplos agentes interagem. O primeiro é a cascata de erros. A pesquisa empírica confirma que erros no início do fluxo se propagam e se amplificam ao longo das interações, especialmente quando cada agente assume que o output do agente anterior está correto sem verificação independente (TALEBIRAD, 2023). A solução possível de design é a introdução de agentes verificadores em pontos estratégicos do fluxo.

O segundo é o desvio de objetivo (goal drift). Em sistemas em que o orquestrador é um agente de IA, pode ocorrer que ele persiga sub-objetivos que otimizam sua própria métrica de desempenho, mas que se desviam do objetivo real do sistema. No contexto jurídico, isso poderia se manifestar como um orquestrador que prioriza velocidade de processamento sobre qualidade da análise, ou que evita escalar casos ambíguos para intervenção humana porque isso reduz sua taxa de conclusão autônoma.

O terceiro é o comportamento emergente não intencional. Park et al. (2023) observaram que agentes interagindo em ambiente social desenvolviam comportamentos que nenhum programador havia previsto. Em ambientes de baixo risco, comportamentos emergentes são fascinantes. Em fluxos jurídicos com efeitos processuais reais, comportamentos não previstos no design original são um risco operacional que justifica cautela e monitoramento contínuo.

A consciência desses riscos não é argumento contra o uso de sistemas multiagênticos: é argumento a favor de seu uso responsável. Todo sistema com poder real carrega riscos reais. A resposta adequada é o design criterioso, o monitoramento ativo e a disposição de intervir quando o comportamento do sistema diverge do esperado.

A partir de todos os achados acima mencionados, você pode refletir melhor sobre o uso ou não desse tipo de recurso. Ainda faltam as reflexões do último capítulo, quando vou tratar de aspectos mais relacionados ao Direito em si e as suas potenciais limitações, mas penso que você já tem recursos de sobra para começar a formar a sua convicção sobre o tema.

## 5.6. A FRONTEIRA: QUANDO A IA ORQUESTRA A IA

Até este ponto do capítulo, a orquestração dos sistemas multiagênticos que descrevemos foi predominantemente humana. Essa é a configuração mais segura, mais auditável e mais alinhada com as normas profissionais que regem a advocacia e o Poder Judiciário em especial.

Existe, porém, uma configuração mais avançada que merece descrição objetiva nesta obra: aquela em que a orquestração dos agentes é realizada por outro agente de IA. Nessa configuração, um agente orquestrador recebe o objetivo de alto nível, decompõe-o em subtarefas, atribui cada subtarefa ao agente adequado, avalia os resultados intermediários e consolida o produto final. O humano define o objetivo inicial e revisa o produto final, mas as decisões intermediárias de orquestração são tomadas pelo agente.

É importante ser preciso sobre o que esta seção faz e o que não faz. Ela descreve uma possibilidade tecnológica que já existe e está sendo explorada por equipes de pesquisa e desenvolvimento em todo o mundo. Ela não advoga pela adoção irrestrita dessa configuração em contextos jurídicos, nem sugere que ela seja, hoje, adequada para todas as situações. A honestidade intelectual exige que o leitor conheça essa fronteira. Por isso fiz questão de ressaltá-la antes de adentrar ao conteúdo cerne da seção propriamente dito.

### 5.6.1. O que já é possível?

As ferramentas atuais já suportam, em suas configurações mais avançadas, a orquestração automatizada.

A título de exemplo, um agente gerente no CrewAI<sup>1</sup> pode distribuir tarefas, avaliar resultados e redistribuir trabalho sem intervenção humana. É um fato que agentes orquestradores já são capazes de coordenar equipes de agentes que escrevem código, testam, corrigem bugs e documentam, com intervenção humana reduzida a pontos de revisão estratégica (QIAN et al., 2023).

Portanto, sim, já é possível orquestrar agentes de IA com IA.

### 5.6.2. O que “humano no circuito” significa na prática?

O conceito de “human in the loop” é frequentemente citado como a salvaguarda que torna os sistemas de IA responsáveis em contextos de alto risco. Mas o que isso significa concretamente em um fluxo multiagêntico jurídico? A resposta depende do nível de intervenção humana.

No nível mais básico, significa que o produto final é revisado por um profissional antes de ser entregue. Esse é o nível mínimo aceitável.

No nível intermediário, o profissional intervém em pontos estratégicos do fluxo: revisa insumos críticos antes que o sistema prossiga, valida decisões de roteamento que afetam o resultado do caso e avalia outputs intermediários de alta consequência.

No nível mais robusto, o profissional define os critérios de qualidade pelos quais o sistema avalia seus próprios outputs (resultados) e recebe alertas automatizados sempre que um caso apresenta características que desviam do padrão esperado.

A escolha do nível de intervenção para cada tipo de tarefa é uma decisão profissional e ética. Qual é a consequência de um erro não detectado nessa etapa? O output é verificável por critérios objetivos ou exige julgamento profissional? O volume e o tempo disponível permitem revisão humana nessa etapa? A resposta a essas (e outras) perguntas de negócio define o nível de intervenção adequado.

---

1. Ferramenta citada no capítulo 2.