

Chamada 2025 de Pesquisadores da Cátedra Oscar Sala

Tema: Governança de Agentes de Inteligência Artificial

A Cátedra, cujo nome presta homenagem ao físico Oscar Sala, patrono da Internet acadêmica no Brasil, faz parte das iniciativas previstas no convênio firmado entre a Universidade de São Paulo (USP) e o Comitê Gestor da Internet (CGI.br) em 21 de agosto de 2020. Tal convênio tem como executantes o Instituto de Estudos Avançados (IEA) e o Comitê Gestor da Internet no Brasil (CGI.br), e seu propósito é fomentar o intercâmbio multidisciplinar para qualificar e ampliar o debate acadêmico e público sobre as tecnologias digitais que podem favorecer o desenvolvimento técnico e científico, a inovação e o direito fundamental de acesso à informação e à comunicação.

Como titular da Cátedra Oscar Sala em 2025, Virgílio Almeida promoverá, ao longo deste novo período, grupos de estudos, seminários e debates com proeminentes pesquisadore(a)s brasileiro(a)s e estrangeiro(a)s sobre temas relacionados à governança dos agentes de inteligência artificial (IA).

Serão discutidos os riscos éticos e sociais relativos à pesquisa e ao desenvolvimento (P&D) dos agentes de IA. Para compreender os desafios desses sistemas autônomos, será adotada uma abordagem multidisciplinar. Além disso, propostas e mecanismos de governança serão desenvolvidos para proteger a sociedade de suas possíveis consequências indesejáveis. As pesquisas deverão procurar identificar os potenciais impactos desses agentes e avaliar soluções que promovam transparência, responsabilidade, visibilidade e alinhamento com valores éticos e sociais fundamentais, pilares-chave para mitigar riscos e assegurar usos éticos e responsáveis dessas tecnologias. Ademais, o contexto tecnológico e social dos países do Sul Global será uma consideração central nas pesquisas sobre os agentes de IA.

Temas

- Autonomia e Escala:** Agentes de IA possuem a capacidade de agir de forma independente, o que aumenta os riscos de erros, mau uso ou consequências não intencionais. Diferentemente dos modelos de linguagem (LLM), suas ações podem impactar diretamente operações sem a necessidade de intervenção humana.
- Interpretabilidade:** As decisões tomadas por agentes de IA frequentemente são pouco transparentes, dificultando o monitoramento, a identificação de falhas e a mitigação de riscos.
- Riscos Éticos:** A autonomia desses agentes pode intensificar desigualdades sociais existentes ou criar tipos de danos, como discriminações, eliminação indiscriminada de postos de trabalho e a perpetuação de preconceitos em decisões críticas.
- Agência:** o *design* e desenvolvimento de agentes de IA devem definir claramente sua autonomia e estabelecer mecanismos eficazes para garantir limites ou salvaguardas (*guardrails*), responsabilidade e prestação de contas (*accountability*).
- Responsabilidade:** é essencial estabelecer marcos legais claros que definam as atribuições de responsabilidade por danos causados por agentes de IA adaptando os princípios tradicionais ao contexto das entidades digitais.
- Visibilidade:** As tecnologias devem assegurar que as ações e os processos decisórios dos agentes de IA sejam transparentes e sujeitos a auditoria.

- Inclusão: O desenvolvimento de agentes de IA deve incorporar perspectivas diversas para ampliar inclusão, reduzir vieses e evitar resultados injustos.
- Arquiteturas tecnológicas de agentes de IA: as arquiteturas referem-se à estruturação dos componentes e processos que permitem o funcionamento de agentes de inteligência artificial. Essas arquiteturas tecnológicas definem como os agentes percebem o ambiente, tomam decisões, planejam ações e interagem com outros sistemas ou usuários. Elas incluem elementos como modelos de linguagem e aprendizado, mecanismos de planejamento, interfaces de comunicação e sistemas de controle, garantindo que os agentes possam executar tarefas de forma autônoma e alinhada aos objetivos definidos.
- Governança adaptativa e colaborativa de agentes de IA: governos, empresas e organizações sociais devem explorar modelos flexíveis de governança dessas tecnologias que acompanham avanços tecnológicos e mudanças regulatórias, promovendo a colaboração entre múltiplos setores (*multistakeholder model*) para alinhar padrões, princípios e compartilhar boas práticas. Além da conformidade imediata, é essencial adotar uma visão de longo prazo, tratando a governança de agentes de IA como uma estratégia para fortalecer a confiança, mitigar riscos e impulsionar a inovação.

O acompanhamento crítico das atividades da cátedra será feito por um grupo composto por até 20 pesquisadore(a)s. O(a)s integrantes se reunirão entre 21/03 a 12/12 de 2025, das 16:00 às 18:00, nas datas informadas, às sextas-feiras, de forma remota ou híbrida. Além dos encontros com convidado(a)s proeminentes, outras atividades serão destinadas aos estudos orientados dos projetos de pesquisa e à produção de textos. Esta chamada não oferece bolsas de estudo ou pesquisa.

Calendário

Datas	Horários	Atividades	Formas de oferecimento
04/04/2025 (6f) 21/03/2025 (6f) [Março]	16:00-18:00	Palestra de abertura do grupo de estudo aberta para o público em geral	Remota
25/04/2025 (6f) [Abril]	16:00-18:00	Reunião do grupo de estudo: definição da bibliografia e dos temas	Remota
27/06/2025 (6f) [Junho]	16:00-18:00	Debate sobre o papel do “ <i>design</i> ” na construção de agentes de IA: em busca de	Remota

		um equilíbrio entre “ <i>human agency</i> ” e “ <i>algorithm agency</i> ”	
18/07/2025 (6f) [Julho]	16:00-18:00	1.a-) Estudo do comportamento de agentes de IA (discussão de estudos empíricos e modelos teóricos)	Remota
15/08/2025 (6f) [Agosto]	16:00-18:00	2.a-) Estudo do comportamento humano nas relações (individuais e coletivas) com agentes de IA	Remota
12/09/2025 (6f) [Setembro]	16:00-18:00	1.b-) Formação de um grupo de estudo e mini-workshop	Remota
26/09/2025 (6f) [Setembro]	16:00-18:00	Debate sobre as formas de colaboração humano-agentes-IA na saúde, trabalho e emprego (discussão e análise de estudos de casos)	Remota
17/10/2025 (6f) [Outubro]	16:00-18:00	2.b-) Formação de um grupo de estudo e mini-workshop	Remota
14/11/2025 (6f) [Novembro]	9:30-17:30	Seminário Internacional sobre o papel dos Agentes de IA e seus impactos sociais, econômicos e políticos nos países	Presencial

		em desenvolvimento com a participação de conferencistas internacionais	
28/11/2025 (6f) [Novembro]	16:00-18:00	Debate sobre o papel dos agentes de IA na sociedade: agentes não humanos devem receber alguma consideração moral e legal como outras entidades que interagem socialmente?	Remota
12/12/2025 [Dezembro]	16:00-18:00	Debate sobre a governança de Agentes de IA e apresentação de propostas e estudos de caso com foco na integração de ciências sociais, governança, direito, saúde, engenharias e computação para gerenciar as mudanças aceleradas pela disseminação de agentes autônomos de IA	Remota

Inscrição

A seleção será feita com base em um único documento de até no máximo duas páginas no qual o(a) pesquisador(a) expresse 1-) seu interesse, 2-) sua motivação, 3 sua contribuição em potencial e 4-) uma curta biografia. As inscrições deverão ser feitas pelo formulário¹ <https://forms.gle/7Pf2trDXa5fchf159> até às 12:00 (meio dia) do dia 10/03/2025.

¹ Para acessar o formulário é preciso realizar login no sistema. Caso não tenha uma conta da plataforma, siga as instruções disponíveis [aqui](#).