
1. Designação da unidade curricular

[4482] Sistemas Generativos e Agentes Inteligentes Multimodais / Generative Systems and Multimodal Intelligent Agents

2. Sigla da área científica em que se insere

INF

3. Duração

Unidade Curricular Semestral

4. Horas de trabalho

162h 00m

5. Horas de contacto

Total: 60h 00m das quais T: 15h 00m | P: 45h 00m

6. % Horas de contacto a distância

Sem horas de contacto à distância

7. ECTS

6

8. Docente responsável e respetiva carga letiva na Unidade Curricular

[760] Arnaldo Joaquim Castro Abrantes | Horas Previstas: N/D

9. Outros docentes e respetivas cargas letivas na unidade curricular

Não existem docentes definidos para esta unidade curricular

10. Objetivos de aprendizagem e a sua compatibilidade com o método de ensino (conhecimentos, aptidões e competências a desenvolver pelos estudantes).

No final da unidade curricular, os estudantes serão capazes de:

1. Compreender e aplicar os fundamentos de modelos generativos (GANs, VAEs, Transformers, Diffusion), reconhecendo o seu impacto em áreas como geração de texto, imagens e outras modalidades.
2. Desenvolver aplicações utilizando grandes modelos de linguagem (LLMs), como GPT, para geração de conteúdo, ajustando modelos pré-treinados para diferentes contextos.
3. Integrar e implementar sistemas multimodais, combinando diferentes tipos de dados (texto, imagem, áudio, vídeo), utilizando modelos como CLIP, BLIP-2, SAM, explorando as suas capacidades em problemas reais.
4. Desenhar e implementar agentes inteligentes, com foco em arquiteturas que integrem LLMs e capacidades multimodais, capazes de realizar tarefas autónomas em ambientes complexos.
5. Analisar criticamente as implicações éticas e sociais da IA generativa e dos agentes inteligentes, desenvolvendo uma consciência dos potenciais impactos e limitações.

10. Intended Learning objectives and their compatibility with the teaching method (knowledge, skills and competences by the students).

By the end of the course, students will be able to:

1. Understand and apply the fundamentals of generative models (GANs, VAEs, Transformers, Diffusion), recognizing their impact on areas such as text, image, and other modality generation.
2. Develop applications using large language models (LLMs), such as GPT, for content generation, fine-tuning pre-trained models for different contexts.
3. Integrate and implement multimodal systems, combining different types of data (text, image, audio, video) using models like CLIP, BLIP- 2, SAM, exploring their capabilities in realworld problems.
4. Design and implement intelligent agents, focusing on architectures that integrate LLMs and multimodal capabilities, capable of performing autonomous tasks in complex environments.
5. Critically analyse the ethical and social implications of generative AI and intelligent agents, developing an awareness of the potential impacts and limitations of these technologies.

11. Conteúdos programáticos

- 1.Introdução aos Modelos Generativos e Multimodais: Definição e aplicações; exemplos como GPT, CLIP e Stable Diffusion; noções de agentes inteligentes.
- 2.Arquitetura Transformers: encoder-decoder, encoder-only (BERT) e decoder-only (GPT); importância do mecanismo de atenção. 3.Modelos Autorregressivos: Fundamentos e aplicações em geração de texto, áudio e imagem; arquiteturas como PixelCNN, WaveNet e GPT.
- 4.Modelos Variacionais: Introdução aos VAEs e suas aplicações.
- 5.Modelos Generativos Adversariais (GANs): Arquitetura de GANs e suas variantes.
- 6.Modelos de Difusão: Princípios e aplicações em geração de imagens com modelos como DDPM e Stable Diffusion.
- 7.Multimodalidade: Integração de múltiplas modalidades (texto, imagem, áudio, vídeo); exploração de modelos como GPT, ViT, CLIP, BLIP-2 e SAM.
- 8.Agentes Inteligentes: Fundamentos e frameworks para desenvolvimento de agentes, com foco em agentes baseados em LLMs e multimodalidade.
- 9.Ética e Impactos Sociais da IA Generativa .

11. Syllabus

- 1.Introduction to Generative and Multimodal Models: Definition and practical applications; examples such as GPT, CLIP, Stable Diffusion; basics of intelligent agents.
- 2.Transformer Architecture: encoder-decoder, encoder-only (BERT), decoder-only (GPT); the importance of the attention mechanism.
- 3.Autoregressive Models: Fundamentals and applications in text, audio, image generation; architectures like PixelCNN, WaveNet, GPT. 4.Variational Models: Introduction to VAEs and their applications.
- 5.Generative Adversarial Models (GANs): GAN architecture and its variants.
- 6.Diffusion Models: Principles and applications in image generation with models like DDPM and Stable Diffusion.
- 7.Multimodality: Integration of multiple modalities (text, image, audio, video); exploration of models like GPT, ViT, CLIP, BLIP-2, SAM.
- 8.Intelligent Agents: Fundamentals and frameworks for agent development, with a focus on agents based on LLMs and multimodality.
- 9.Ethics and Social Impacts of Generative AI.

12. Demonstração da coerência dos conteúdos programáticos com os objetivos de aprendizagem da unidade curricular

Os conteúdos programáticos estão alinhados com o objetivo de oferecer uma formação abrangente nos fundamentos dos modelos generativos e suas aplicações, promovendo uma abordagem crítica e responsável em relação às tecnologias emergentes. Isso proporciona aos alunos as competências técnicas necessárias para atuar na área da engenharia informática. A introdução aos modelos generativos e multimodais expõe os estudantes às capacidades, limitações e riscos dessas tecnologias. O estudo das arquiteturas Transformers e do mecanismo de atenção fundamenta a compreensão dos processos de geração multimodal. A inclusão de modelos autorregressivos, variacionais, GANs e de difusão oferece uma visão diversificada das técnicas disponíveis, capacitando os alunos a aplicar esses conhecimentos em cenários reais. Por fim, o desenvolvimento de agentes inteligentes equipa os alunos com ferramentas para participar ativamente na construção do futuro nesta área.

12. Evidence of the syllabus coherence with the curricular unit's intended learning outcomes

The program content aligns with the goal of providing comprehensive training in the fundamentals of generative models and their applications, fostering a critical and responsible approach to emerging technologies. This equips students with the necessary technical skills to work in the field of computer engineering. The introduction to generative and multimodal models exposes students to the capabilities, limitations, and risks of these technologies. The study of Transformer architectures and the attention mechanism lays the groundwork for understanding multimodal generation processes. The inclusion of autoregressive, variational, GAN, and diffusion models offers a diverse view of available techniques, enabling students to apply this knowledge in real-world scenarios. Finally, the development of intelligent agents equips students with the tools to actively contribute to shaping the future in this field.

13. Metodologias de ensino e de aprendizagem específicas da unidade curricular articuladas com o modelo pedagógico

A unidade curricular adota uma abordagem pedagógica top-down, começando pela exploração de aplicações práticas de IA e evoluindo gradualmente para a compreensão dos conceitos técnicos subjacentes. O ensino combina aulas teóricas, onde são introduzidos os conceitos fundamentais, e aulas práticas, nas quais os estudantes trabalharão com sistemas generativos e multimodais, implementando pequenos modelos utilizando o PyTorch como framework base. O Google Colab será utilizado para experimentar modelos, proporcionando um ambiente acessível e poderoso sem a necessidade de instalação de software nos computadores pessoais dos alunos, que podem ter capacidades variadas.

As metodologias incluem demonstrações guiadas, exercícios práticos em laboratório e trabalhos em grupo, permitindo que os estudantes apliquem os conhecimentos adquiridos na criação de modelos e agentes inteligentes. A utilização de plataformas como Hugging Face para aceder a modelos pré-treinados, juntamente com ferramentas como Streamlit ou Gradio para desenvolver aplicações interativas, fomentará o trabalho colaborativo e a experimentação. Este método assegura que os alunos desenvolvam competências técnicas e capacidade de análise crítica, alinhando-se aos objetivos de aprendizagem estabelecidos.

13. Teaching and learning methodologies specific to the curricular unit articulated with the pedagogical model

The course adopts a top-down pedagogical approach, starting with the exploration of practical AI applications and gradually progressing to the understanding of the underlying technical concepts. The teaching combines theoretical classes, where fundamental concepts are introduced, with practical classes, in which students will work with generative and multimodal systems, implementing small models using PyTorch as the main framework. Google Colab will be used for model experimentation, providing an accessible and powerful environment without the need for software installation on students' personal computers, which may have varying capabilities.

The methodologies include guided demonstrations, practical lab exercises, and group work, allowing students to apply their acquired knowledge in creating models and intelligent agents. The use of platforms such as Hugging Face to access pre-trained models, along with tools like Streamlit or Gradio for developing interactive applications, will foster collaborative work and experimentation. This method ensures that students develop both technical skills and critical analysis abilities, aligning with the established learning objectives.

14. Avaliação

Avaliação distribuída com exame final.

A avaliação desta unidade curricular é contínua e baseada na aplicação prática dos conceitos estudados. A classificação final será obtida através da média ponderada dos seguintes elementos de avaliação:

1. Exame (30%): Este componente avaliará a compreensão teórica dos conceitos abordados ao longo do semestre. Para aprovação, os alunos devem obter uma nota mínima de 9.5 neste exame.
2. Trabalhos Práticos (30%): Ao longo do semestre, os estudantes desenvolverão pequenos projetos focados na implementação de modelos generativos, multimodais e de agentes inteligentes. Estes trabalhos testarão a capacidade de aplicar frameworks como PyTorch e Hugging Face para resolver problemas reais, com uma nota mínima de 9.5 necessária para a componente prática.
3. Projeto Final com Discussão (40%): No final da unidade, os estudantes realizarão um projeto mais aprofundado, integrando os conhecimentos adquiridos para desenvolver uma aplicação multimodal ou um sistema de agentes inteligentes. A avaliação incluirá a qualidade técnica, inovação e a capacidade de apresentar os resultados.

A participação nas discussões e nas atividades práticas será valorizada, incentivando o envolvimento ativo e a troca de ideias. Esta avaliação garante uma progressão prática e teórica, promovendo o desenvolvimento de competências técnicas e críticas. Para obter aprovação, o aluno deverá ter uma média final igual ou superior a 9.5 (escala 0-20) em todos os componentes de avaliação.

Todas as componentes são pedagogicamente fundamentais.

14. Assessment

Distributed assessment with final exam.

The evaluation of this course is continuous and based on the practical application of the concepts studied. The final grade will be obtained through the weighted average of the following assessment components:

1. Exam (30%): This component will assess the theoretical understanding of the concepts covered throughout the semester. A minimum grade of 9.5 is required to pass the exam.
2. Practical Assignments (30%): Throughout the semester, students will develop small projects focused on the implementation of generative, multimodal models, and intelligent agents. These assignments will test the ability to apply frameworks such as PyTorch and Hugging Face to solve real-world problems, with a minimum grade of 9.5 required for the practical component.
3. Final Project with Discussion (40%): At the end of the course, students will undertake a more in-depth project, integrating the knowledge acquired to develop a multimodal application or an intelligent agent system. The evaluation will include technical quality, innovation, and the ability to present results.

Participation in discussions and practical activities will be valued, encouraging active involvement and exchange of ideas. This assessment ensures a balance between practical and theoretical progress, promoting the development of both technical and critical thinking skills. To pass the course, students must achieve a final average grade of 9.5 (on a 0-20 scale) in all assessment components.

All the components are pedagogically fundamental.

15. Demonstração da coerência das metodologias de ensino com os objetivos de aprendizagem da unidade curricular

As metodologias de ensino e avaliação estão estruturadas para garantir que os estudantes desenvolvam tanto uma compreensão teórica sólida quanto competências práticas em sistemas generativos, multimodais e agentes inteligentes. A abordagem top-down facilita uma aprendizagem progressiva, começando com a utilização de modelos pré-treinados em ambientes práticos e avançando para a construção de modelos próprios. Aulas teóricas, seguidas de atividades práticas com ferramentas como PyTorch e Hugging Face, promovem a integração do conhecimento com a experimentação.

A avaliação contínua, através de trabalhos práticos e um projeto final, permite aos alunos aplicar os conceitos em problemas reais, assegurando que desenvolvem competências técnicas avançadas e a capacidade de resolver problemas complexos. No entanto, para garantir que a compreensão teórica está devidamente consolidada, a avaliação inclui também um exame escrito, representando 30% da nota final. O exame assegura que os alunos dominam os fundamentos teóricos essenciais que sustentam os modelos e técnicas, complementando a avaliação prática. A participação ativa nas discussões e atividades em sala de aula reforça a capacidade crítica e de comunicação, alinhando-se com o objetivo de preparar os estudantes para o uso aplicado e crítico da IA generativa e multimodal.

15. Evidence of the teaching methodologies coherence with the curricular unit's intended learning outcomes

The teaching and assessment methodologies are structured to ensure that students develop both a solid theoretical understanding and practical skills in generative, multimodal systems and intelligent agents. The top-down approach facilitates progressive learning, starting with the use of pre-trained models in practical environments and advancing to the construction of their own models. Theoretical lectures, followed by practical activities using tools such as PyTorch and Hugging Face, promote the integration of knowledge with experimentation.

Continuous assessment, through practical assignments and a final project, allows students to apply concepts to real-world problems, ensuring they develop advanced technical skills and the ability to solve complex issues. However, to ensure that theoretical understanding is properly consolidated, the assessment also includes a written exam, representing 30% of the final grade. The exam ensures that students grasp the essential theoretical foundations underlying the models and techniques, complementing the practical evaluation. Active participation in discussions and class activities reinforces critical thinking and communication skills, aligning with the goal of preparing students for the applied and critical use of generative and multimodal AI.

16. Bibliografia de consulta/existência obrigatória

- [1] Simon J. D. Prince (2023), Understanding Deep Learning, MIT Press
- [2] Jay Alamar, Maarten Grootendorst (2024), Hands-On Large Language Models, O'Reilly Brown,
- [3] T. et al. (2020). Language Models are Few-Shot Learners. NeurIPS.
- [4] Ramesh, A. et al. (2021). Zero-Shot Text-to-Image Generation. OpenAI.
- [5] Kirillov, A. et al. (2023). Segment Anything. Meta AI.
- [6] Li, J. et al. (2023). BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models, Salesforce Research.
- [7] Yin, S. et. al. (2024). A Survey on Multimodal Large Language Models, IEEE TPAMI.

17. Observações

Unidade Curricular Obrigatória

Data de aprovação em CTC:

Data de aprovação em CP: