

PROGRAMA DE CURSO

Código	Nombre			
CC66V	Inteligencia Artificial Generativa			
Nombre en Inglés				
Generative Artificial Intelligence				
SCT	Créditos	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
	3	24	0	72
Requisitos			Carácter del Curso	
Propósito del Curso				
Al finalizar este curso, se espera que los/las estudiantes sean capaces de: (i) comprender y aplicar técnicas avanzadas de inteligencia artificial generativa para la creación y manipulación de diversos tipos de contenido, incluyendo texto, imágenes y audio, (ii) entender los principios detrás de los modelos generativos más utilizados en la actualidad, como redes generativas adversarias (GANs), modelos de difusión y modelos de lenguaje de gran escala (LLMs y LMMs), y (iii) utilizar bibliotecas y herramientas especializadas tipo <code>transformers</code>, o <code>peft</code>, para la implementación eficiente de estas técnicas en distintos contextos profesionales. El curso abordará tanto los fundamentos como los avances recientes en IA generativa, incluyendo técnicas de optimización y eficiencia en LLMs (como el <code>parameter efficient fine-tuning</code> y la cuantización), generación de audio, y modelos multimodales. Se prestará especial atención a la personalización y mejora de modelos a través de enfoques como Retrieval-Augmented Generation (RAG) y el uso de agentes basados en IA. Durante el curso, los y las estudiantes trabajarán con casos de estudio que reflejan desafíos típicos en la aplicación de IA generativa, lo que les permitirá desarrollar habilidades en la modelización, implementación y ajuste de estos modelos. Asimismo, aprenderán a evaluar la calidad de los contenidos generados y a abordar problemas comunes como el sesgo en modelos generativos y la alucinación en LLMs. Finalmente, se espera que los y las estudiantes adquieran una comprensión profunda de las implicaciones éticas del uso de modelos generativos, incluyendo su impacto en la sociedad y las mejores prácticas para un desarrollo y aplicación responsables de esta tecnología en distintos ámbitos profesionales.				
Resultados de Aprendizaje				
Al finalizar este curso, se espera que los y las estudiantes sean capaces de: 1. Comprender y aplicar técnicas avanzadas de inteligencia artificial generativa para la creación y manipulación de diversos tipos de contenido, incluyendo texto, imágenes y audio.				

2. Entender los principios detrás de los modelos generativos más utilizados en la actualidad, como las redes generativas adversarias (GANs), los modelos de difusión y los modelos de lenguaje de gran escala (LLMs y LMMs).
3. Utilizar bibliotecas y herramientas especializadas, como transformers o peft, para implementar estas técnicas de manera eficiente en distintos contextos profesionales.

Metodología Docente	Evaluación General
La metodología consiste en clases teóricas, revisión de artículos de investigación, trabajos prácticos y un proyecto.	50% tareas, 30% trabajo práctico y 20% cuestionarios.

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
Clase N° 1	Introducción y Modelos de Visión Generativos	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- El Generative Modeling - GANs, Variational Autoencoders, Modelos de difusiones - Conditional GANs, Stable Diffusion, StyleGANs	Comprender los fundamentos teóricos de los modelos generativos para visión por computadora y sus aplicaciones en generación de imágenes sintéticas.	[1, 2, 3]

Número	Nombre de la Unidad	Duración en Semanas
Clase N° 2	Modelos de Lenguajes Largos (LLM)	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
LLM: - Language Modeling - Basics: BERT, GPT-2, T5 - Abilities and In-Context-Learning - Tokenization and Training - Instructions and Alignments	Dominar los conceptos fundamentales de modelos de lenguaje de gran escala, sus capacidades emergentes y técnicas de optimización.	[4, 5]

- Reasonings Training in Practice - Evaluating LLMs LLMs as Agents Aplicación: - Cuantización - Adapters y Parameter Efficient Fine Tuning		
--	--	--

Número	Nombre de la Unidad	Duración en Semanas	
Clase N° 3	Efficiency and Deployment of Generative Models	1	
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía	
- DPO, RLHF, and Alignment - Techniques for efficiency (quantization, LoRA, pruning) - Deployment strategies for generative AI - Model optimization for production	Aplicar técnicas de optimización y despliegue de modelos generativos en entornos de producción, considerando eficiencia computacional y alineamiento.	[4, 5]	

Número	Nombre de la Unidad	Duración en Semanas	
Clase N° 4/5	Modelos multimodales	2	
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía	
Multimodalidad: - Introducción y conceptos básicos - Fusion and Coordination - Contrastive Learning and other losses	Entender los fundamentos de modelos que procesan múltiples modalidades de datos simultáneamente. Implementar y evaluar modelos multimodales de gran escala para tareas de comprensión y generación de contenido.	[6]	

Large Multimodal Models: - LXMERT, ViBERT, ViLT - CLIP, BLIP, xGen-MM - Stable Diffusion - Flamingo and Frozen - Open Source training dataset and Evaluation - Video - Tokenization		
--	--	--

Número	Nombre de la Unidad	Duración en Semanas	
Clase N° 6	Modelos de Audio (no solo generativos)	1	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- Audio synthesis (WaveNet, Tacotron) - Speech and music generation - Modelos largos de Audio - HuBERT - Jukebox - Whisper		Desarrollar aplicaciones de IA procesando el audio, incluyendo generación de voz y música.	[6,7]

Número	Nombre de la Unidad	Duración en Semanas	
Clase N° 7	Reasoning and Advanced Prompting Techniques	1	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía

- Advanced prompt engineering - Enhancing reasoning capabilities in LLMs - Chain-of-thought prompting and complex task solving - ReAct	Dominar técnicas avanzadas de prompt engineering y métodos para mejorar las capacidades de razonamiento de los modelos de lenguaje.	[4,5,8]
---	--	----------------

Número	Nombre de la Unidad	Duración en Semanas
Clase N° 8	Advanced LLMs: Agents, Tools, and RAG	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- Building intelligent agents - Integrating external tools and APIs - Retrieval-Augmented Generation (RAG) for complex tasks	Construir sistemas inteligentes que integren modelos de lenguaje con herramientas externas y bases de conocimiento.	[4,5,8]



Referencias Bibliográficas

- [1] David Foster, Generative Deep Learning, 2nd Edition. " O'Reilly Media, Inc."
- [2] Sanseviero, O., Cuenca, P., Passos, A., & Whitaker, J. (2024). Hands-On Generative AI with Transformers and Diffusion Models. " O'Reilly Media, Inc."
- [3] HuggingFace's Diffusion Models Course <https://huggingface.co/learn/diffusion-course>
- [4] Tunstall, L., Von Werra, L., & Wolf, T. (2022). Natural language processing with transformers. " O'Reilly Media, Inc."
- [5] Alammar, J., & Grootendorst, M. (2024). Hands-on large language models: language understanding and generation. " O'Reilly Media, Inc."
- [6] CMU-MMML: <https://cmu-mmml.github.io/>
- [7] HuggingFace's AI Audio <https://huggingface.co/learn/audio-course/>
- [8] HuggingFace's AI Agents Course <https://huggingface.co/learn/agents-course/>

Vigencia desde:	16/07/2025
Elaborado por:	Valentin Barriere, Andres Abeliuk, Clemente Henriquez

**Draft Clem:**

Desarrollar competencias para comprender, implementar y aplicar técnicas avanzadas de inteligencia artificial generativa en la creación y manipulación de diversos tipos de contenido digital, incluyendo texto, imágenes, audio y contenido multimodal, utilizando modelos generativos modernos y bibliotecas especializadas para su implementación.



Libros de Texto Fundamentales

- [1] Deep Learning. Ian Goodfellow, Yoshua Bengio, Aaron Courville. 2016. ISBN: 9780262035613.
<http://www.deeplearningbook.org/>
- [2] Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. David Foster. O'Reilly Media, 2019.
- [3] Natural Language Processing with Transformers. Lewis Tunstall, Leandro von Werra, Thomas Wolf. O'Reilly Media, 2022.

Papers Fundamentales - Modelos Generativos de Visión

- [4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).
<https://arxiv.org/abs/1406.2661>
- [5] Kingma, D. P., Welling, M. "Auto-Encoding Variational Bayes." ICLR (2014). <https://arxiv.org/abs/1312.6114>
- [6] Ho, J., Jain, A., Abbeel, P. "Denoising Diffusion Probabilistic Models." NeurIPS (2020).
<https://arxiv.org/abs/2006.11239>

Papers Fundamentales - Transformers y LLMs

- [7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. "Attention is all you need." Advances in neural information processing systems 30 (2017).
<https://arxiv.org/abs/1706.03762>
- [8] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL-HLT (2019). <https://arxiv.org/abs/1810.04805>
- [9] Brown, T., Mann, B., Ryder, N., et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020). <https://arxiv.org/abs/2005.14165>
- [10] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. "Improving Language Understanding by Generative Pre-Training." OpenAI (2018). https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf
- [11] Raffel, C., Shazeer, N., Roberts, A., et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." Journal of Machine Learning Research 21.140 (2020): 1-67. <https://arxiv.org/abs/1910.10683>

Papers de Eficiencia y Despliegue

- [12] Hu, E. J., Shen, Y., Wallis, P., et al. "LoRA: Low-Rank Adaptation of Large Language Models." ICLR (2022). <https://arxiv.org/abs/2106.09685>
- [13] Ouyang, L., Wu, J., Jiang, X., et al. "Training language models to follow instructions with human feedback." Advances in Neural Information Processing Systems 35 (2022). <https://arxiv.org/abs/2203.02155>
- [14] Bai, Y., Jones, A., Ndousse, K., et al. "Constitutional AI: Harmlessness from AI Feedback." arXiv preprint (2022). <https://arxiv.org/abs/2212.08073>



[15] Rafailov, R., Sharma, A., Mitchell, E., et al. "Direct Preference Optimization: Your Language Model is Secretly a Reward Model." arXiv preprint (2023). <https://arxiv.org/abs/2305.18290>

[16] Dettmers, T., Pagnoni, A., Holtzman, A., Zettlemoyer, L. "QLoRA: Efficient Finetuning of Quantized LLMs." arXiv preprint (2023). <https://arxiv.org/abs/2305.14314>

Papers de Modelos Multimodales

[17] Radford, A., Kim, J. W., Hallacy, C., et al. "Learning transferable visual representations with natural language supervision." International conference on machine learning (2021). <https://arxiv.org/abs/2103.00020>

[18] Li, J., Li, D., Xiong, C., Hoi, S. "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding." International Conference on Machine Learning (2022). <https://arxiv.org/abs/2201.12086>

[19] Jia, C., Yang, Y., Xia, Y., et al. "Scaling up visual and vision-language representation learning with noisy text supervision." International Conference on Machine Learning (2021). <https://arxiv.org/abs/2102.05918>

[20] Baltrušaitis, T., Ahuja, C., Morency, L. P. "Multimodal machine learning: A survey and taxonomy." IEEE transactions on pattern analysis and machine intelligence 41.2 (2018): 423-443. <https://arxiv.org/abs/1705.09406>

Papers de Large Multimodal Models

[21] Alayrac, J. B., Donahue, J., Luc, P., et al. "Flamingo: a visual language model for few-shot learning." Advances in Neural Information Processing Systems 35 (2022). <https://arxiv.org/abs/2204.14198>

[22] Ramesh, A., Dhariwal, P., Nichol, A., et al. "Hierarchical text-conditional image generation with clip latents." arXiv preprint (2022). <https://arxiv.org/abs/2204.06125>

[23] Liu, H., Li, C., Wu, Q., Lee, Y. J. "Visual instruction tuning." Advances in neural information processing systems 36 (2024). <https://arxiv.org/abs/2304.08485>

[24] Li, J., Li, D., Savarese, S., Hoi, S. "BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models." International Conference on Machine Learning (2023). <https://arxiv.org/abs/2301.12597>

Papers de Modelos de Audio

[25] van den Oord, A., Dieleman, S., Zen, H., et al. "WaveNet: A generative model for raw audio." arXiv preprint (2016). <https://arxiv.org/abs/1609.03499>

[26] Shen, J., Pang, R., Weiss, R. J., et al. "Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions." 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). <https://arxiv.org/abs/1712.05884>

[27] Dhariwal, P., Jun, H., Payne, C., et al. "Jukebox: A generative model for music." arXiv preprint (2020). <https://arxiv.org/abs/2005.00341>

[28] Radford, A., Kim, J. W., Xu, T., et al. "Robust speech recognition via large-scale weak supervision." International Conference on Machine Learning (2023). <https://arxiv.org/abs/2212.04356>



[29] Hsu, W. N., Bolte, B., Tsai, Y. H. H., et al. "HuBERT: Self-supervised speech representation learning by masked prediction of hidden units." IEEE/ACM Transactions on Audio, Speech, and Language Processing 29 (2021): 3451-3460. <https://arxiv.org/abs/2106.07447>

Papers de Reasoning y Prompting

[30] Wei, J., Wang, X., Schuurmans, D., et al. "Chain-of-thought prompting elicits reasoning in large language models." Advances in Neural Information Processing Systems 35 (2022): 24824-24837. <https://arxiv.org/abs/2201.11903>

[31] Wang, X., Wei, J., Schuurmans, D., et al. "Self-consistency improves chain of thought reasoning in language models." arXiv preprint (2022). <https://arxiv.org/abs/2203.11171>

[32] Yao, S., Zhao, J., Yu, D., et al. "ReAct: Synergizing reasoning and acting in language models." arXiv preprint (2022). <https://arxiv.org/abs/2210.03629>

[33] Yao, S., Yu, D., Zhao, J., et al. "Tree of thoughts: Deliberate problem solving with large language models." Advances in Neural Information Processing Systems 36 (2024). <https://arxiv.org/abs/2305.10601>

Papers de Agents, Tools y RAG

[34] Lewis, P., Perez, E., Piktus, A., et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in neural information processing systems 33 (2020): 9459-9474. <https://arxiv.org/abs/2005.11401>

[36] Schick, T., Dwivedi-Yu, J., Dessì, R., et al. "Toolformer: Language models can teach themselves to use tools." arXiv preprint (2023). <https://arxiv.org/abs/2302.04761>

[37] Park, J. S., O'Brien, J. C., Cai, C. J., et al. "Generative agents: Interactive simulacra of human behavior." Proceedings of the 36th annual acm symposium on user interface software and technology (2023). <https://arxiv.org/abs/2304.03442>

[38] Zhao, W. X., Zhou, K., Li, J., et al. "A survey of large language models." arXiv preprint (2023). <https://arxiv.org/abs/2303.18223>