

Human-Centered Generative AI for Content Generation

Hao-Wen Dong

University of Michigan
hwdong@umich.edu

Generative AI has been transforming the way we interact with technology and consume content. The recent successes of LLM-based chatbots, AI assistants and text-to-image generation systems have showcased how AI-powered technology can augment human creativity and boost human productivity. In the next decade, generative AI technology will also reshape how we create music, audio and video content in entertainment, commercial and educational fields, including the music, film, TV, podcast and gaming sectors.

Despite the successes of generative AI in certain fields, it has been challenging to integrate generative AI into the professional creative workflow for content creation for several reasons: 1) new application domains may require new generative models; 2) professionals need assistive tools that augment their creativity and productivity in addition to fully automated tools; 3) certain media require handling multimodal data streams at the same time.

My research aims to address these challenges and **augment human creativity with machine learning**. I develop human-centered generative AI technology that can be integrated into the professional creative workflow, with a focus on music, audio and video content creation. In this proposed New Faculty Highlights talk, I will survey my representative work on three main themes of my research: 1) novel generative models for new domains, 2) AI-assisted tools for content creation and 3) multimodal generative models for content creation.

Novel generative models for new domains First, I will describe how I develop novel generative models for new domains that aims at enabling deep learning in underexplored application fields. This part will highlight the following work:

- My work published on generating multi-instrument music using generative adversarial nets, which represents the first deep learning model for multitrack polyphonic music generation (Dong et al., AAAI 2018), as well as follow-up work (Dong and Yang, ISMIR 2018; Dong et al., ISMIR 2020, ICASSP 2023; Xu et al., AIMG 2023; Ryu et al., ISMIR 2024)
- Our ongoing work on generating teasers for long documentaries, which has opened up a new pathway towards long-range multimodal generative models for videos (Xu et al., arXiv 2024)

AI-assisted tools for content creation Second, I will talk about how I build AI-assisted content creation tools that aim to augment human creativity in their creative workflow. I will highlight the following work:

- My work on automatic music instrumentation that can produce convincing instrumentation for a solo piece (Dong et al., ISMIR 2021)
- My work on synthesizing expressive violin performance from sheet music (Dong et al., ICASSP 2022; Kim, Dong, and Jeong, arXiv 2024)

Multimodal generative models for content creation Third, I will introduce my work on multimodal generative models for content creation that can process, understand and generate data in multiple modalities at the same time. I will cover the following work in this part:

- Our ongoing work on text-to-symbolic music generation that aims at generating sheet music from natural language descriptions (Xu et al., ISMIR LBD 2024, arXiv 2024)
- My work on self-supervised multimodal models that learn to separate and synthesize sounds from watching noisy videos for text-queried sound separation (Dong et al., ICLR 2023) and text-to-audio synthesis (Dong et al., WASPAA 2023)

Finally, I will close the talk by discussing my future research plans on 1) multimodal generative AI for content creation, 2) human-AI co-creative tools for music, audio and video creation, and 3) human-like machine learning algorithms for music, movies and arts. My long-term goal is to lower the barrier of entry for content creation and democratize professional content creation for everyone.

Intended and Expected Audience

- Researchers, policy makers and industry representatives interested in knowing how generative AI can be applied to creative and interdisciplinary fields
- Researchers working in close fields such as multimodal learning, natural language processing, speech processing and computer vision
- Students working on or interested in applied AI research in music, audio and video content creation