Recent advances in text-to-music (TTM) and music source separation systems enable new forms of human–AI collaboration in creative production workflows. However, their practical integration into real-world music–making contexts remains understudied. To investigate how such models can support professional creativity, we developed an interactive demo used in a study with music producers, who were invited to explore and test AI models as creative production aids. The goal was to assess how AI tools can be meaningfully incorporated into existing workflows, evaluate their controllability, and observe emerging creative strategies. We now propose this demo for the conference to facilitate broader discussion on AI-assisted music creation.

Models and framework selection. For music generation we selected **MusicGen**, a text- or melody-conditioned autoregressive model available in three sizes, and for source separation we used **HT-Demucs**, which relies on dual U-Nets and a Transformer architecture. We adopted the *htdemucs_6s* variant, capable of separating vocals, drums, bass, and other stems, and specifically providing separation for VDBO, guitar, and piano in our setup. The combination of MusicGen with HT-Demucs allows for independent manipulation of generated musical sources, enhancing compositional flexibility. While the recently released MusicGen-Stem integrates generation and separation into a unified model, it became available only after the beginning of our study and currently supports only three stems. Furthermore, although recent works introduce extended control mechanisms for TTM models, text remains the most widely adopted and intuitive input modality. Given the study's limited time frame and our focus on understanding producers' interactions with text-based workflows, we did not incorporate MusicGen's melody conditioning.

Interface design. The interface was designed following the AI controllability exploration principles proposed in [40]. Taking inspiration from prior interactive systems, we implemented multiple-output generation options alongside general and domain-specific controls, enabling users to explore diverse musical outcomes and better understand model behaviour. To determine which controls to include, we reviewed interfaces publicly available on Hugging Face. Given the constraints of the interactive study—limited session duration, need for ease of use, and minimal learning overhead, we opted for a streamlined design with essential controls only. The interface features two tabs: one for music generation and one for source separation. In the generation tab, users can choose output duration, model size, and text prompt. The separation tab provides stems for drums, bass, guitar, piano, and other instruments, excluding vocals, as MusicGen does not generate vocal tracks.

The demo was developed using **Gradio**, chosen for its intuitive layout, rapid prototyping capabilities, and strong support for machine learning models. A custom interface layer was implemented to ensure full experimental control: one author developed the interface, and two authors thoroughly tested it. The demo was created specifically for the interactive user study, in which music producers engaged with the system to evaluate how AI tools can be incorporated into real creative processes.

Conclusion. This work presents an integrated system combining state-of-the-art music generation and source-separation models with a controllable and user-friendly interface purpose-built for studying human–AI creative workflows. By proposing this demo for the conference, we aim to foster discussions on practical deployment, user-centered design considerations, and the evolving role of AI in audio and music production and creation.

Hardware needed: our laptop