Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 19 No. 8 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijet.v19i08.50223

SHORT PAPER

# Building Consistent Characters through Open-Source Generative AI

**Luca Martini(✉), Lara La Tessa, Daniele Zolezzi**

University of Genoa, Genoa, Italy

luca.martini@edu.unige.it

**ABSTRACT**

This paper presents a pipeline for creating consistent characters using open-source generative AI tools, aimed at enhancing educational content. Through the use of *XTTS*, *Fooocus*, and *FaceFusion*, a talking avatar of the marine biologist Raffaele Issel was developed, accurately replicating his facial and vocal features. This study demonstrates how these tools can generate high-quality, engaging educational materials with minimal technical expertise and resources. The methodology facilitates the creation of interactive avatars for use in graphic novels and comic books, enriching museum and educational experiences.

**KEYWORDS**

generative AI (genAI), consistent characters, open source

## 1 INTRODUCTION

Generative AI (genAI) tools allow for the rapid production of high-quality contents, providing educators with a valuable resource to create engaging learning experiences that maintain students' interest and attention [1]. GenAI is revolutionizing the field of character creation, offering solutions that streamline processes and improve consistency across various applications. One notable use of genAI is in 2D gaming, where it is transforming character design and animation. These tools enable developers to manage animation randomness, greatly reducing the time and complexity involved in creating 2D game animations [2]. Beyond gaming, genAI is utilized to craft hyper-realistic characters that mimic human appearances and behaviors. These AI-generated characters serve beneficial roles in diverse fields such as entertainment, education, and therapy, offering new ways to engage and educate users. This underscores the vast potential of these technologies to impact various industries by producing characters that are both authentic-looking and -sounding [3].

Consistency in character traits is key in storytelling and virtue ethics, enabling a coherent narrative and helping readers connect with characters. Consistent characters behave in predictable ways, fostering empathy and allowing readers to anticipate

Martini, L., La Tessa, L., Zolezzi, D. (2024). Building Consistent Characters through Open-Source Generative AI. *International Journal of Emerging Technologies in Learning (iJET)*, 19(8), pp. 82–89. https://doi.org/10.3991/ijet.v19i08.50223

Article submitted 2024-05-22. Revision uploaded 2024-07-11. Final acceptance 2024-07-11.

their actions. In literature, characters with stable traits and behavior throughout the story are seen as more believable and relatable. Research shows that consistent characters in both verbal and non-verbal cues tend to have a stronger influence on people's behavior and are generally preferred by participants [4]. The integration of genAI in character creation represents not merely a technological advancement but a gateway to new creative possibilities, increased productivity, and innovative tools for artistic and educational expression. These applications highlight the potential of AI to not only emulate human creativity but also to expand it in exciting new directions.

This paper aims to demonstrate a possible pipeline for creating consistent characters using open-source tools that generate images, spoken audio, and videos for educational purposes. The use of digital and interactive tools, such as mobile apps, augmented reality, and virtual reality, can enrich the museum experience and engage audiences more effectively. However, implementing these technologies requires substantial investments in hardware and software, often exceeding the budget constraints of many museums. This paper presents a pipeline built entirely on open-source systems, capable of creating consistent characters in real-world settings and integrating them into educational projects. After outlining the workflow, a case study will be discussed to demonstrate the results achieved using this methodology.

## 2 OUR PIPELINE FOR CREATING CONSISTENT CHARACTERS

A foundational approach to developing consistent characters leverages open-source genAI frameworks, which manage the complete workflow from designing avatars' facial features to generating speech videos. These frameworks are accessible via platforms such as *Hugging Face* and can be utilized through the *Pinokio* software. *Pinokio* serves as an open-source AI application gateway that streamlines the management of AI applications, models, or workflows using one-click scripts, facilitating backend operations such as installations and configurations.

The initial phase of an avatar development process involves choosing a voice synthesis system that can precisely emulate the spoken content. *XTTS* effectively facilitates voice cloning across multiple languages with significant ease and minimal data requirements. By requiring only a six second audio sample as "*Reference Audio*" and a "*Text Prompt*" to read, *XTTS* minimizes the need for extensive training datasets, thereby efficiently speeding up the voice cloning process. *XTTS* advances voice synthesis by introducing features that improve accessibility and allow for extensive customization. It supports multilingual capabilities, enabling voice generation and cloning in 17 different languages, which enhances its global use. The technology simplifies the cloning process, requiring only a short audio sample and allowing the transfer of emotional tones and styles, thereby enriching the expressive quality of voices. Additionally, *XTTS* facilitates the creation of multilingual content through its cross-language cloning feature and ensures high-quality voice reproduction with a 24 kHz sampling rate. This tool can be employed by content creators to construct dialogues that will later be articulated by characters. However, the system is optimized for short sentences. Consequently, despite being completely free, the system may experience a slight slowdown in processing particularly lengthy speeches.

The subsequent phase entails enhanced control over the scenes where our character will be positioned. To introduce additional details to our character, we employ *Fooocus*, an image generation software derived from the *Gradio* Python library. This tool is valuable due to its effective image control system, which enables the provision of detailed information on facial features and body positioning. It utilizes

existing images in the "*ImagePrompt*" as a source to extract necessary information for replicating these details in the final output of the generation process. "*FaceSwap*" function allows users to define specific facial features for the character under development, while "*PyraCanny*" function aids in determining the precise placement of the character within the image. Before initiating image generation, it is imperative to review and adjust the settings. The result, however, from the tests conducted generates a character that perfectly maintains the position determined by "*PyraCanny*" but fails to retain the precise facial features of the initial character.

To address this issue, an additional tool, *FaceFusion*, an open-source solution, can be employed. *FaceFusion* not only has the capability to replace a face in an image with that of our character to maintain consistent identity, but it also refines facial details in images, enhancing clarity and providing a more polished appearance. A notable functionality of *FaceFusion* is its ability to synchronize lip movements in videos with an audio track, revolutionizing character interaction with audiences. To utilize this feature, one should select "*lip_syncer*" rather than "*face_swapper*" within the "*Frame Processors*" menu, then upload the desired audio track as the character's voice and specify the target video for lip synchronization. For optimal results, it is recommended to keep the "*occlusion*" option active within "*Face Mask Types*" and enable "*face_enhancer*" in "*Frame Processors*." A crucial step in this process is converting the image into a video that matches the duration of the audio track, ensuring perfect lip synchronization throughout the video, thus bringing the character to life with exceptional realism.

## 3 CASE STUDY

A recent research has found and demonstrated the effectiveness in educational and museum contexts of storytelling articulated through the creation of virtual avatars using artificial intelligence [5], [6]. These virtual avatars can be deployed to narrate historical events and periods and biographies of distinguished figures from history through their faces and voices for an interactive and engaging experience.

Following this line of inquiry, the effectiveness of the proposed pipeline is evaluated by creating a talking avatar of Raffaele Issel, a prominent marine biologist at the University of Genoa who lived during the late 19th and early 20th centuries. The objective is to develop an avatar that recounts the life and work of Issel, highlighting his significant contributions to marine biology through his dedication and research.

Most of the information about Raffaele Issel comes to us through the works of Cantucci [7], Conci [8], and the latest contribution by Alessandro Pellerano [9]. His life is marked by intense research activity and great academic successes, making Raffaele Issel a cornerstone of the history of the University of Genoa and marine biology. Raffaele Issel (1878–1936) earned his degree in natural sciences from the University of Genoa in 1900, with a thesis on Italy's thermal fauna, and soon exhibited a keen interest in marine biology and hydrobiology. After studying at the Zoological Station in Naples and the Oceanographic Museum of Monaco, Issel became an assistant in zoology in Modena and later a professor of marine biology in Genoa, where he co-founded a marine laboratory at Quarto dei Mille in 1911 [10]. He significantly advanced marine biology in Italy, focusing on marine ecology, thermal fauna, and plankton. His work integrated modern methodologies with ecological studies of marine organisms, and he was active in various scientific societies and oceanographic research.

The aim of this case study is to develop images of Raffaele Issel using open-source genAI tools, which can later be used for educational purposes to inform readers about his life and works. The generated images can be utilized to create in-depth videos about the biologist and can also be used in graphic novels or comic books, making the topics covered more engaging and captivating. This study seeks to demonstrate how individuals can generate consistent character appearances with minimal effort and without extensive knowledge of photomontage tools. The capabilities offered by these tools are significant, even for organizations that cannot afford extended timelines and high costs associated with complex projects such as comic book creation. These tools have the potential to make educational and museum experiences more engaging and memorable.

# 4 RESULTS OBTAINED

## 4.1 *XTTS*

The first step in creating our character is to give him a voice capable of narrating the story of his life and discoveries. The idea is to animate Raffaele Issel, having him explain everything related to his marine biology research within the interactive pages of the comic book. This study examines an introductory mini-video designed to welcome users to the Quarto dei Mille laboratory. To achieve this, the previously mentioned *XTTS* was provided with an eleven-second "*Reference Audio*" of a man with a Genoese accent, compensating for the absence of original audio from the scholar. The provided prompt, shown in Figure 1, resulted in audio that is consistent with both the input text and the initial "*Reference Audio*." This audio will subsequently be used to create the final video of our character, where lip sync will be synchronized with the images generated as described in the subsections.

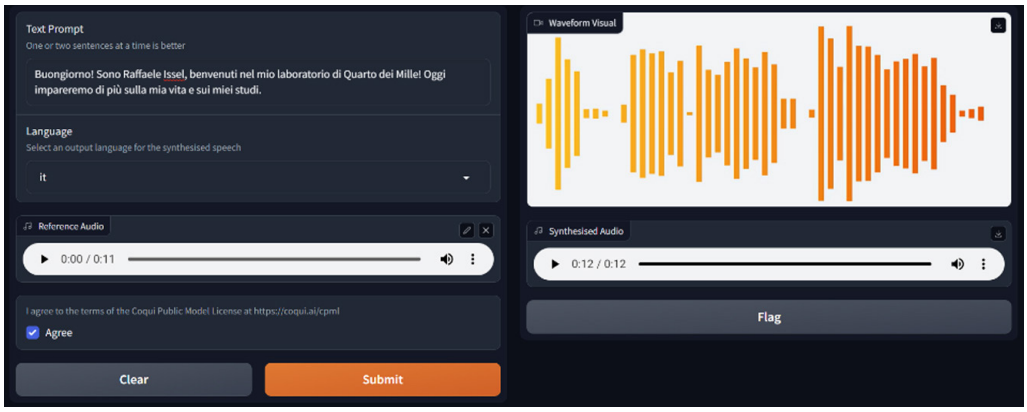


**Fig. 1.** *XTTS* prompt for the realization of Raffaele Issel's voice

## 4.2 *Fooocus*

The method outlined in our pipeline involves using *Fooocus* to create images that incorporate Raffaele Issel's facial features, utilizing existing images of this historical figure. This approach allows us to accurately depict the subject's characteristics without the need to generate a face from scratch. Within the program, we select the "*Input Image*" and "*Advanced*" options to access the program's supplementary capabilities and choose to work through "*ImagePrompt*."

The first image we provide will be a snapshot of the interior of the Quarto dei Mille laboratory, setting the "*ImagePrompt*" command to provide contextual information about the scenario where the created character will be placed. The second image is a photograph of Raffaele Issel, which will provide the basic facial information that the program will use to create the character through the "*FaceSwap*" command (see Figure 2).

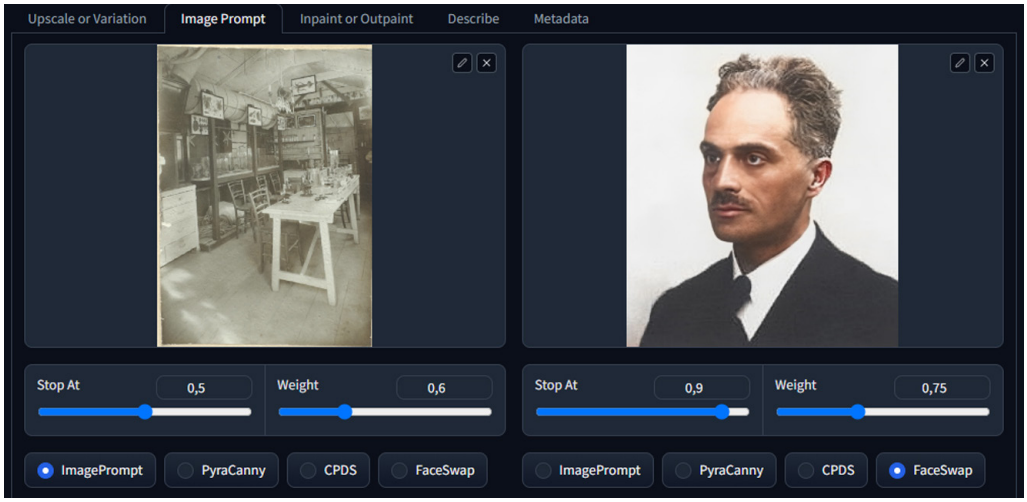


**Fig. 2.** Images supplied for *Fooocus* system

We apply the preferred preset to the photos, in our case "*Default*," and select the "*Speed*" performance type to optimize between the quality of the created image and the speed of execution. We choose the preferred "*Aspect Ratios*," in our case 1024 × 1024, the number of images to be generated, and the format in which they will be produced in the right column within the "*Settings*." If there are elements that you absolutely need to avoid in your image, remember to add a "*Negative Prompt*" in the same settings column. The prompt given to create the new image was as follows: "*Create an image of a middle-aged man in his laboratory, sitting at a desk. He is dressed in early 20th-century clothing. The image should be with HDR (High Dynamic Range) effects, emphasizing the intricate details and textures of the scene. The laboratory should have a vintage look, with scientific equipment and papers scattered on the desk, giving a realistic and immersive feel of the period.*"

The results obtained were good from the very first attempts, as shown in Figure 3. Both the background of the Quarto dei Mille laboratory and the biologist's face remained faithful to what we proposed in the "*ImagePrompt*." Each element proposed is consistent with what is required, and the context given by the prompt is respected. However, to achieve even greater fidelity to the original face, we will process the newly created images through additional open-source software to make them even closer to the original.

**Fig. 3.** Results obtained using the procedure described in Section 4.2

### 4.3 *FaceFusion*

The final step to perfect our character is to use *FaceFusion*. The results from the previous step were already very similar to the original subject, but they were not perfectly identical, which could reduce the credibility and coherence of the scenes to be created. To further enhance the facial detail, we provide *FaceFusion* with the image of Raffaele Issel that we already used in *Fooocus* to provide the basic details of our character. We select "*face_swapper*" for face replacement and "*face_enhancer*" to improve the final result quality as the "*Frame Processors*". The "*Target*" images to be modified are the same ones generated in step 4.2. The results obtained, visible in Figure 4, show that the character now maintains a consistent face across all images and is ready to be used for creating new content.


**Fig. 4.** Comparison of the results obtained after using *FaceFusion*

The images obtained after using the "*face_swapper*" now need to be transformed into a video with a duration at least equal to the audio for which we want to create the video. Once the base video is obtained using the preferred video editing software, we return to FaceFusion and select "*lip_syncer*" for the "*Frame Processors*" to perform lip sync between the video and the audio, while keeping "*face_enhancer*" to further enhance the facial quality in the final result. Additionally, based on our tests, to achieve more realistic results, it is recommended to select only the "*occlusion*" option under "*Face Mask Types*" for a more accurate representation of the speaking character. The final result can be viewed through the QR code in Figure 5.

**Fig. 5.** QR code to see the talking avatar of Raffaele Issel

## 5 CONCLUSIONS

The implementation of the described pipeline has demonstrated the effectiveness of open-source genAI tools in creating consistent characters for educational purposes. Using *XTTS*, *Fooocus*, and *FaceFusion*, we successfully developed a talking avatar of the marine biologist Raffaele Issel, accurately replicating his facial and vocal features. This methodology enables the generation of engaging, high-quality educational content without requiring advanced technical skills or significant financial resources.

The results indicate that the combination of these tools can be effectively used to create characters that enhance learning experiences in both museum and educational contexts. The ability to produce talking avatars makes historical narratives more vivid and interactive, increasing audience engagement and comprehension.

Future work should focus on integrating the generated avatars into augmented reality (AR) and virtual reality (VR) environments, creating immersive experiences for museums and schools. Additionally, expanding the dataset of reference images and audio to include other historical figures will improve the variety and quality of the content. Furthermore, these tools can be employed to create graphic novels and comic books featuring historical figures such as Raffaele Issel, making the topics more engaging and accessible. These advancements will strengthen the role of genAI in education and cultural dissemination, opening new frontiers for interactive and immersive learning experiences.

## 6 REFERENCES


[1] D. Zolezzi, S. Iacono, and G. V. Vercelli, "Star words re-generated: Gamification and GenAI for effective training," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 19, no. 4, pp. 97–104, 2024. https://doi.org/10.3991/ijet.v19i04.47977

[2] S. Qiu, "Generative AI processes for 2D platformer game character design and animation," *Lecture Notes in Education Psychology and Public Media*, vol. 29, pp. 146–160, 2023. https://doi.org/10.54254/2753-7048/29/20231440

[3] V. Danry *et al.*, "AI-generated characters: Putting deepfakes to good use," in *CHI EA 22 Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, no. 19, pp. 1–5. https://doi.org/10.1145/3491101.3503736

[4] K. Isbister and C. Nass, "Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics," *Int. J. Hum. Comput. Stud.*, vol. 53, no. 2, pp. 251–267, 2000. https://doi.org/10.1006/ijhc.2000.0368

[5] P. Pataranutaporn *et al.*, "AI-generated characters for supporting personalized learning and well-being," *Nature Machine Intelligence*, vol. 3, pp. 1013–1022, 2021. https://doi.org/10.1038/s42256-021-00417-9

[6] P. Pataranutaporn *et al.*, "Living memories: Ai-generated characters as digital mementos," in *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 2023, pp. 889–901. https://doi.org/10.1145/3581641.3584065

[7] R. Cantucci, "La vita e le opere di Raffaele Issel," *Atti della Società di scienze e lettere di Genova*, pp. 19–40, 1936.

[8] C. Conci, "Repertorio delle biografie e bibliografie degli scrittori e cultori italiani di entomologia," in *Memorie della Società entomologica italiana.* Pagano, 1975. [Online]. Available: https://books.google.it/books?id=mqUZzwEACAAJ

[9] A. Pellerano, "L'istituto di zoologia della universita'degli studi di genova-cenni storici e ricordi," in *BMIB-Bollettino dei Musei e degli Istituti Biologici*, vol. 75, 2013.
[10] R. Issel, "Il piccolo laboratorio marino di Quarto dei Mille," *Bios*, vol. 1, nos. 2–3, 1914.

## 7 AUTHORS

**Luca Martini** is a PhD Student in Digital Humanities at the University of Genoa, Italy. His research focuses on Generative AI, VR, and 3D rendering (E-mail: luca.martini@edu.unige.it).

**Lara La Tessa** is a digital humanist and a PhD student at the University of Genoa, Italy. Her research focuses on innovative approaches to digitizing and enhancing the public collections of the University of Genoa, incorporating User Experience Design and Gamification methodologies, with a touch of Artificial Intelligence (E-mail: lara.latessa@edu.unige.it).

**Daniele Zolezzi** is a PhD Student in Digital Humanities at the University of Genoa, Italy. His research focuses on Gamification, Serious Games, Virtual Reality, and Generative Artificial Intelligence (E-mail: daniele.zolezzi@edu.unige.it).