AIMCAT : 'I Feel the Wires'

Tomas Nilhén, **AIMCAT team** (https://aimcatmusic.com/team2021/), #AIMCAT, **Catalonia**
Website: https://aimcatmusic.com | Twitter: @AImCAT2021 | Instagram: aimcat_2021 | Contact: tomas.nihlen@gmail.com

Abstract— **AIMCAT is a Catalan team and project that was born from a tweet by a Swede. We have become a platform for constant dialogue, given the diversity and the plurality of the group. In fact, the different AI tools have turned into full members of the team. We must admit that human and AI co-creation was really complex at the beginning, but let the result speak for itself: a song called 'I feel the wires'.**

Keywords—AI Song Contest, Artificial Intelligence, Music, Audiovisual, Co-creation, Ethical & Social considerations

I. AIMCAT TEAM

One of the most amazing, even crazy, things about this project was its conception. It all started when Tomas Nilhén, a Swede living in Catalonia sent a tweet. Tomas is a rising tweet star in Catalonia mainly because he is learning and practicing Catalan on twitter, but also because, in doing so, he talks about common people's daily problems and routines. As a software engineer interested in art and creativity, he got into the loop of the AI song contest [1]. A few weeks before the kick-off of the 2021 edition, he sent that very tweet that we all will keep in our memories forever.

The answer to his call on Twitter was well beyond his expectations, and everything started to move fast and forward. Software engineers, writers, AI evangelists, philosophers, visual artists, 3D animators, song writers, machine learning engineers, psychologists, composers, singers and journalists, all sorts of profiles popped up. Today, our Slack workspace is crammed with +33K messages from the 60 people that have access to it. No one knew each other and the only link between us was that we all were Tomas followers on Twitter. At the beginning there was some chaos, but from that chaos emerged the word and the music and the visuals and the communication strategy and the ethics project. Now, six weeks later, we are about to find an association for the promotion of artificial intelligence in artistic projects.

At the end of the XIXth Century (1899), a Swiss came to Catalonia and founded a football club that ended up being the Barcelona FC. You never know if history is going to be repeated, this time round, with a daring Swede ...

II. DIALOGUE: THE STORYTELLING OF THE SONG

As shown below, we are a very diverse group of people with different backgrounds and different degrees of understanding of the technical nuances of AI technology. People with a technical background and deep AI knowledge tend to be more enthusiastic about the technology in part because they know what the current limitations are and how to tackle de risks. People less acquainted with the technology tend to have more mixed feelings of excitement and fear, or just scepticism.

This is why, from the beginning, the AIMCAT team was very clear that our mantra was 'dialogue' in all its forms: at the time of creating multidisciplinary teams; within the dynamics of the teams; in the creation process itself, intertwining human part and artificial intelligence; using a plurality of AI tools. Consequently, the storytelling of the song is centered around 'dialogue'. Thus, the global message of

the song, that will be highlighted in the chorus: 'Blending (love/respect) leads to evolution'.

Dramaturgy, the song as a story: There is a dialog (not a love story) between human and AI based on admiration and respect. We wanted to avoid crossed readings about gender, sex, romanticism, or cliché... The message of the song is in a different layer, a more humanistic one. For example: -Human: 'I would like to think like you do.'- AI: 'I would like to have a heart beat like you have'.

Structure of the song: The structure of the song is divided into 10 sections. Each of the sections has an internal coherence between music and lyrics and all together, as a sequence, they have a narrative structure.

Section 1 Intro	A chord progression generated by AI moved us to use the first 16 bars to open a 'sound' gate to a different music proposal. We used classical sounds and electronic percussion. The story of the whole theme is that the evolutionary force of the universe moves dualities to dialog and move forward, creating new entities or 'beings'. AI and humans communicate and use each other to improve. The Vitruvian Man by Da Vinci is the starting point.
Section 2 Chorus	This chorus carries the basic dramaturgy lines: Duality, dialog, fear, joy, evolution. The drum beat of the snare drum in the first and third verses is out of the downbeat to create a sense of precaution, while in the second and fourth goes downbeat to generate a sense of progress.
Section 3 Stanzas 1 and 2	The many stops of the first stanza and the final progression pretend to show the feeling of a slow and caution forward movement. A silence section of full stop in the second stanza shows the 'lapse' or 'void' humans fear, as stated in the chorus lyrics. A strings progression (strings taken as the evolution force of the universe) shows that the 'progress is solid', 'beautiful', and often seems 'chaotic'.
Section 4 Chorus	This chorus shows that the two entities are losing their fears and want to get closer.
Section 5 Stanza 3	The arrangements are somewhat chaotic, because the two entities are moving towards the other. This step guides them to the 'interface' (the Reactable) that will place them in a new "world" or scenario.
Section 6 Chorus	In this chorus, the two entities play with the 'interface' and learn how to know each other better. The fear is gone. They play freely.
Section 7 Digression	The orchestra sounds mean the evolutionary forces of the universe. The singers and the lyrics here are 100% AI generated. The message: 'It's wonderful to be alive' The strings go smooth to show relief, but there is a backing rhythm of tension, because in every evolution there is some drama.
Section 8 Step back chorus	Then, there is the fear of being out of the comfort zone, but still the joy of the experience.
Section 9 Let's move chorus	Here the AI generated voices appear again with the melody of the digression and blend with the melody of the chorus. New voices appear, too high pitched and low pitched. Everything is allowed, with joy.
Section 10 Ending	We are back to the beginning. The 'sound' gate closes, and the new Vitruvian paradigm is now a woman, a blending of both entities: AI+Human. The chord progression here was generated by AI.

Finally, four versions of the song have been created: completely in English, completely in Catalan, in English with the two final choruses in Catalan, and Catalan with the two final choruses in English. We cast a vote among all the AIMCAT team members to decide which version should go to the contest. Dialogue in capital letters. The English version with the chorus in Catalan won.

III. ORGANISATION

The organisation and operation of the work teams has been dynamic at all times. In fact, teams have been created and shut down as work for the ultimate goal was needed. We even changed 'team leaders', when it was best for AIMCAT. The working and communication tool has been Slack and plenary meetings were held with all teams once a week on Sundays via Zoom. At the end of the process AIMCAT is made up of 35 people plus our souls in AI.

A. Teams and coordination among teams

The teams of the whole song creation process have been intertwined in a dialogue with the AI strategy team and the team that developed the software and the corresponding tools. There is also a series of teams that wrapped the matrix: Social Impact & Ethics, Group Dynamics and Communication. And all together we formed the general plenary of the AIMCAT. The leader of the orchestra was the coordination group led by the Project Manager (PM). See Figure 1.
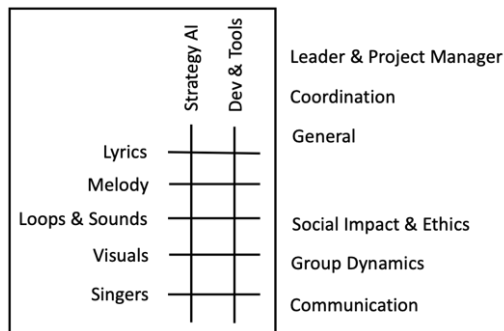

Figure 1. Teams & coordination among teams.

For specific needs, working groups have been created with very specific tasks such as: creating stanzas, initial tests, music references, coordinating the IDEAL-recording space, or writing this process document.

B. Critical moments

Obviously, there have been critical moments, but the creativity and enthusiasm of the team has managed to resolve them and make us stronger. As an example, we will cite two of them:

- At one point, the person leading the team and acting as project manager (PM) believed that it was best to transform the workgroups and pass the baton to another PM who wanted to do it 'collegially'. And here we are.
- Panic ensued when one of the singers fell ill 24 hours before recording the voices. So not only did we find and enrol another singer, but she also did very well.

We feel proud of our work as a team: everybody has contributed from their role to keep the team strong, just as the 'Castellers' do when build their human castles.

IV. CREATIVE PROCESS AND USE OF AI TOOLS

In the same way that we decided that the plot line of the song and the video should be a dialogue, we considered that co-creation would also be a dialogue between human creation and the creation that came from artificial intelligence. Therefore, it has been an intertwined process.

The field work has emerged as one of the key parts of the creation process: there have been many contributions that may not be in the final version, but without them the result would not have been possible: both human contributions (initial lyrics written by writers), and some AI-based algorithms that have not been fully exploited.

In the following, we will detail this dialogue with AI in lyric, melody, music, voice and visual.

A. Lyrics

We have approached the writing of the lyrics as a computer assisted community creation process. In other words, we explored the frontier between artificial intelligence and crowd intelligence. The step-by-step process in a simplified version was for the English text:

1. Build stanzas of four verses by randomly matching two people that wrote one verse at a time alternately.

2. Use the OpenAI API to feed GPT-3 with one stanza of four verses and ask it to generate a new one.

3. Human cherry picking of the results produced by the model to build the song with blocks of 4 verses both for chorus and stanza.

The interaction with the model was done through a Python script which incorporated a filter on the output of the model. This filter was adjusting the number of verses per stanza and the minimum and maximum length of the verses. Although some of the words that appear in the song were in the human versions that were fed into the model, all the verses of the song are outputs of the model.

We tried to replicate this process with the Catalan version of the song. However, even if the Catalan outputs from GPT-3 were really amazing in most cases, they were not consistent enough to build the whole song. Thus, another approach was taken:

1. A canonical translation by a member of the team was made and given as a reference to the other team members.

2. The rest of the team was dropping free versions of the translation in Catalan.

3. At the end, 8 different versions were transferred to the music team to be used in the final composition. They used verses from different proposals and also made better versions to be fit into the melody.

Table 1 shows the final chorus in English and Catalan.

English Chorus version	Catalan Chorus version
I feel the wires, I feel the wires My blood flows and my heart beats I fear the lapse, I fear the lapse I keep discovering new things	*Em sento els fils, em sento els fils i el meu cor com batega. Tinc por del buit, tinc por del buit, però no paro d'aprendre.*

Table 1. Final Chorus in English and Catalan.

B. Melody

We used AI at different points in the melody:

a) Melody Generation

We propose a Genetic Algorithm (GA) to melody generation. GA is an algorithm inspired by natural selection in which a population mixes and evolves to improve. The genetic algorithm has been tuned to support musical components (keys, bars, scales, among others) and users had a Google Colab interface to set each feature.

Finally, we mixed different sounds to create melodies around the song. On top of a "melody accompaniment", we created a "melody generation" (mixing dorian and lydian musical modes), that matched well with the spirit of the song. Furthermore, we generated a synthetic voice to sing those intertwined melodies (see subsection D).

Unfortunately, for a matter of time, the GA was not incorporated into the final version.

b) Melody Continuation

We structured the basic chord progression of the song around a melody given by the melody continuation tool. Around a simple melody of a few notes, we heard a really attractive melody that helped us build the core chords of the song. After that, we followed the composition work around that basic structure (Dm-Gm-Bb-G). So, we built the song structure around this idea given by the AI. The AI-tools used are:

- Magenta Music Transformer. A Transformer model that continues a melody given by the user. It uses the given melody as a "primer" and auto-completes the rest, similar to how GPT-3 operates on words.

- Magenta RNN. A Recurrent Neural Network trained on MIDI that, given a melody, will continue it.

c) Melody Accompaniment

For the Melody Accompaniment we use:

- Magenta Music Transformer. A Transformer model that produces accompaniment chords / bassline for a melody given by the user.

The intro/outro chords were created with such technique. We gave the tool 4 notes, and it created the harmonic accompaniment. Sounded unusual, and we liked it as the initial theme in the introduction of the song.

C. Music

For the Music Interpolation we also use Magenta MusicVAE (Variational Auto-Encoder): A probabilistic auto-encoder model that, given two MIDI files representing two musical parts (consisting of melody, bass and drums), will project them to latent space and interpolate between them, producing a smooth transition between the two parts.

The motto of the project was 'Blending leads to evolution' and this is how we approached the use of AI tools described above. The learning process also had an impact on the final result, mimicking the storyline of the song.

From the very beginning the process started by introducing simple melodies to the algorithms and playing with the output to build the intro (section 1). This was like setting the tone for the structure of the whole song. As we were scattered among different locations, we interacted mainly remotely, and MIDI files were traveling back and forth. Quite often people's faces showed astonishment as the proposals of the AI were really pushing things beyond the comfort zone of musicians.

As musicians' skills on using AI tools improved, there was an increasing coherence between the human proposals and the algorithms' outputs. So, by the end of the iterative phase there was material to build up to 10 songs. The final step was to ensemble the instrumental version with all the hues that appeared along the process.

The production of the song where vocals, synthetic voices and music were put together for the final song that we present in the competition was performed with the standard means.

D. Synthetic Voice

Technologies around synthetic voice have witnessed great progress in the last years thanks to the interest in personal assistants. However, singing voices represent a major challenge that is far from yielding results at any point comparable with human voices.

We have used synthetic voices in post-production of the song. The singing voice was generated using the library midi2voice. Midi2voice is basically a way to interact with the sinsy.jp website from the Nagoya Institute of Technology which implements a HMM-based Singing Voice Synthesis System.

Sung synthetic voices are one of the main AI challenges. By manually adjusting the input text for fitting into the score and adjusting the spelling to improve English pronunciation the results were acceptable but far from having the quality to be used in a competition. The great idea that changed the team perception on synthetic voices was to make it sing in different tones. So, we generated four voice layers that, combined, recreated the voice effect which reminds that of vocoder, a classic tool of sound processing, just that we have done it 100% digitally with AI algorithms and text embeddings.

E. Visuals

We mainly approached three phases:

1. *Planning, Script and Story Board*: We focused first on the main idea (dialogue between AI and a human) then we transcript from the Script to an Story Board.

2. *Shooting sessions*: We had shooting sessions in two different stages, one of them is the IDEAL space in Barcelona, a 2000m2 immersive projection room where we projected a complete motion graphics base for the dancers, using a REACTABLE system in the middle of the room as a communication tool (fiction) between AI and a human.

3. *Editing the Music Video*: After this we edited all the shots, adding the necessary VFX in a Music Video format.

Different visual elements have been incorporated into the final production of the video. We would like to highlight some that show the will to maintain dialogue with AI and technology even in the small details.

a) Model 3D

DeepMotion's Animate 3D employs various state-of-the-art techniques from deep neural networks to bio-mechanical simulations in order to reconstruct the best possible 3D motion perceived from 2D video. Their AI analyzes each frame of the input video, identifies the various joints of the humanoid figure, reconstructs the root trajectory of the figure and the ground planes, and then translates them into 3D space, enabling markerless motion capture.

This tool permitted us to modify its look to our liking later. This AI is responsible for analysing a video that we can upload and track a complete humanoid skeleton, and then pass it to the Model and end up wearing it with lights or particles.

b) Others but not less important

For enlarging the resolution of the final shot, we used the services of www.deep-image.ai, and also to add some extra effects to the dancers, we used EbSynth App.

Among many other visual references and putting in connection different elements of reference (the 5ths circle in music, Ramon Llull drawings on computation, and Joan Miro's Art) we created the "starbyte", a cypher-code that expresses 8 binary positions, in

schemes of 8 bits, and within a star of 8 spikes. We integrated this artistic development in the visual field of the video, as a metaphor of dialogue between humans and machines, used to enrich the visual result.

c) Reactable

For the filming of the video, we used the image of a Reactable provided by a UPF research group. The Reactable is a new electronic musical instrument with a simple and intuitive design, which enables musicians to experiment with sound, change its structure, control its parameters and be creative in a direct, refreshing and unseen way.

F. Summarising

To conclude this section, we show two summary tables. Table 2 shows the interaction between AI and the human process and Table 3 summarizes all the models and techniques used in each music building block.

	Lyrics	Melody Generation	Melody Continuation	Melody Accompaniment	Loops & Sounds	Music	Visuals	Singers		
Process	1	1	1	1	1	1	1	■		
Output	3	5	4	2	2	4	1	1		
Legend	1	Some Human / Some AI								
	2	Human content								
	3	AI conditioned generation (pipeline)								
	4	AI generated, human curated content								
	5	AI generated together (jointly)								
	■	Part did not exist								

Table 2. Interaction between AI and the human process.

Music building blocks	Models & Techniques
Lyrics	GPT-3
Melody Generation	Genetic Algorithm
Melody Continuation	Transformer, RNN
Melody Accompaniment	Transformer
Music	Variational Auto-Encoder
Visuals	DeepMotion's Animate 3D (Deep learning), Cypher-code own, Reactable (as a visual resource)
Singers (Synthetic voice)	HMM-based Singing Voice Synthesis System

Table 3. Models and techniques used in each music building block.

V. DIVERSITY AND ETHICAL CONSIDERATIONS

From the very beginning, it became clear that there had to be a group that followed the process of creating the song, keeping the most ethical and social aspects in mind. We have even been in contact with the Observatory of Ethics in Artificial Intelligence of Catalonia (OEIAC). In order to assess these aspects, two surveys were carried out on all AIMCAT members: one focused on Social Impact & Ethics and the other on assessing how group dynamics had worked.

Social Impact & Ethics. The survey was answered by 27 people. From the resulting report we could highlight that 96.3% of the members perceive the team as diverse. This matches well the goal set by the leader, which he explained in his answer to the survey, of building a diverse group that would guarantee different points of view and we have succeeded!

Despite the efforts, the representation of the male gender has ended up being the majority, however, the presence of women in this project (33.3%) exceeds by 15 points the average of women specialists in the field of technology in the EU (17.7 %) according to the Women in Digital Scoreboard 2020.

Group dynamics. The survey was answered by 21 people. From the final report we could highlight the main values: cooperation (70%), respect (90%) and intergroup communication (90%). As a line of improvement more information is needed on the goals to be achieved and resources allocated to do so.

Colorblind. The project has also taken other considerations into account. For example, the color palette chosen in the visual part is adapted to the colorblind collective using the Adobe Color accessibility tool. It is a mathematically 'square' palette (four complementary colors) which is 'safe palette for colorblind people'.

Finally, we consider that AIMCAT promotes six of the *Sustainable Development Goals* (SDG): 3, 5, 8, 9, 10 and 17.

VI. FINAL DISCUSSION

If we were to start over now, we would probably be better acquainted with the complexity of intertwining artificial intelligence into the whole song process and move faster. However, we would not change the value brought on by working and sharing with a team of people like AIMCAT. It has been a pleasure and an honour to share and meet such valuable people.

In fact, one should remember that this project has been carried out without a budget, and with the sole contribution of the personal time and means of many professionals. And each of them has been instrumental in the final result. This also holds true for the communications team, which is now intensively beginning its task of making us known and getting votes from June 1. That's why the authorship shall be shared by all members of the team equally.

VII. CONCLUSIONS AND THOUGHTS FOR THE FUTURE

Conclusions. This adventure has managed to deliver a song that we are very happy with, both for the process and the result. Obviously, there is always room for improvement and innovation.

As we commented at the beginning of this document, after this enriching experience we want to go one step further by creating an association that gives continuity to the values of AIMCAT by promoting the dialogue between artificial intelligence and art in Catalonia.

Thoughts for the future. We are very pleased that the deadline for the video is set for just before the start of the voting. This should be maintained. It would also be good to have a few more pages in the process document.

Enjoy our song!

REFERENCES

[1] Cheng-Zhi Anna Huang, Hendrik Vincent Koops, Ed Newton-Rex, Monica Dinculescu, Carrie J. Cai, "AI Song Contest: Human-AI Co-Creation in Songwriting", ISMIR 2020, Cite as: arXiv:2010.05388