

Diversity and Representation in ICCC: A review of Computational Creativity publication content and authorship trends 2017-2022

Marzia Bilwani¹ and Anna Jordanous²

¹ Independent Researcher, Pakistan. marziabilwani@gmail.com

² School of Computing, University of Kent, Canterbury, Kent, UK. a.k.jordanous@kent.ac.uk

Abstract

This paper examines issues of diversity and representation in the International Conference on Computational Creativity (ICCC), evaluating the diversity of authors, content and existing creative systems. We consider: the diversity of cultural context of published systems; diversity of authors in terms of gender and culture; how the pandemic affected diversity; and trends in content of accepted papers. The study covers a period of six consecutive years around the COVID-19 pandemic (2017-22), to better understand the impact of the pandemic on diversity. The research team includes individuals from different career stages and research interests, who bring diverse perspectives to the analysis. We evidence the need for greater diversity in both authorship and content, as well as in the creative systems discussed in the proceedings. The paper concludes with recommendations for increasing diversity and representation in the field of computational creativity.

Introduction

As the world becomes more globalized and interconnected, the importance of maintaining diversity increases. Thus, in the last two decades, there has been significant discourse on ensuring diversity and representation. For example, in 2019, the percentage of top-grossing films having female protagonists more than doubled from 16 percent in 2002 to 40 percent in 2019 (Lauzen 2020). However, diversity and inclusion issues abound in academic conferences (Walters 2018).

In this paper we focus on points specific to the International Conference of Computational Creativity (ICCC) in the past six years. We consider: the diversity of cultural context of published systems; diversity of authors in terms of gender and culture; how the pandemic affected diversity; and trends in content of accepted papers. Our time period of six consecutive years since 2017 helps us consider impact of the COVID-19 pandemic on diversity and representation during the years around COVID-19 worldwide lockdowns.

This paper's team of computational creativity researchers includes people with ICCC experience ranging from none (prior to 2023) to several years' engagement with ICCC; the team spans different continents and different career points.

Importance of Diversity There are many benefits of fostering diversity. For example, in the workplace, cultural diversity has been found to lead to process gains by enhancing satisfaction and creativity (Stahl et al. 2010). The Association of Computational Creativity Task Force conducted a valuable recent study of diversity in terms of demographics of people involved in ICCC (Cunha et al. 2020). Loughran (2022) discussed biases arising due to data: systems learn and create from the data they have been trained on. She also reflected upon detrimental biases relating to demographics, particularly in the representation of women in ICCC.

The importance of diversity in the creative fields stems largely from the ability of art to preserve culture and allow individuals a form of self-expression. While the goals of computational creativity may be different from those of art, by creating systems that can enhance human creativity or are capable of being creative, it can be implied that they are indirectly fulfilling the same purpose.

Diversity becomes vital, as AI systems may tend to contribute to lesser diversity in the 'collective experiments of life' and lead to greater standardization in decision-making (Loi, Vigano, and van der Plas 2020). Computational Creativity (CC) essentially challenges this standardisation, for a unique way to curb negative societal effects of AI systems.

Analysis of existing systems

In the last six years, a total of 344 papers were published. To analyse creative systems discussed in these papers, categorisation was done by reviewing each paper manually. Categorisations in conference proceedings have changed over the years. So here we focused on classifying those papers reporting artefacts being produced or systems aiding creative production, as: Cuisine, Music, Text, Visual, Coding, Sound, Problem-solving, Dance, Theatre, Games or Other.

When analysing the systems, the top three categories that emerged were producing text-based artefacts, followed by visuals and then music. This was a trend seen across all six years with 2017 and 2021 being the exceptions. The high number of systems producing text-based artefacts could be attributed to how this category included systems generating stories, headlines, poetry, song lyrics and jokes.

Year	Num papers	Num authors	Mean authors per paper (3sf)	Num countries (continents)
2017	<i>34</i>	93	2.74	14 (4)
2018	46	<i>91</i>	<i>1.98</i>	16 (3)
2019	59	140	2.37	22 (5)
2020	85	199	2.34	24 (4)
2021	65	171	2.63	30 (5)
2022	55	148	2.69	<i>16 (4)</i>

Table 1: Numbers of papers in proceedings per year for ICCC’17-22, and corresponding number of authors involved and number of countries (and continents) represented. Highest values are in bold font, lowest in italics.

Text Within this category, for most cases, the artefacts being produced were found to be in English. A few exceptions existed with systems generating French poetry (Hämäläinen, Alnajjar, and Poibeau 2022) and Portuguese headlines (Mendes and Oliveira 2020). However, the number of these systems is quite a small proportion of the overall number of papers in this category (10 out of 75 papers produce artefacts in languages other than English). However, all of these languages have European origins. One system studies Japanese popular entertainment narratives (Murai et al. 2022), but with generation of plot analyses rather than text.

Visual Arts For systems producing visual artefacts, it was observed that where systems were producing paintings it was mostly the generation of Western styles using the WikiArt dataset. Japanese-influenced art is the notable exception, with ICCC publications including a Japanese facial art database (Tian et al. 2020), Ukiyo-e stylistic generation (Tian et al. 2021) and Manga graphic novel generation (Melistas et al. 2021). Availability of datasets plays a significant part in affecting the diversity of art produced using AI algorithms (Loughran 2022; Burgdorf et al. 2022). A search for datasets using the term ‘art’ on Kaggle and Google (Hillier 2022) showed that the top ten results consist of artworks produced by western artists. WikiArt website contains art from 106 countries worldwide, but has 89 countries not represented (wikiart.com 2023).

Music Genres across systems producing musical artefacts tended to be limited to a few categories such as classic, jazz and pop. American and British songs, or just English songs were a popular choice (Harris, Harris, and Bodily 2020; Gordon et al. 2022). Out of 28 systems generating music, only three were found to produce Italian, Spanish and German music (Ackerman, Morgan, and Cassion 2018; Navarro and Oliveira 2018; Banar and Colton). One limitation of this evaluation, however, was that not all papers provided information on the dataset being used. Some systems were also producing instrumental music and it becomes difficult to evaluate how culturally diverse they are.

(Preliminary) gender analysis

The gender of the authors submitting a paper is also an important aspect influencing how diverse or creative systems are. Therefore, gender data was analysed for 149 people who have published in more than one year in the period under study. This excludes 467 people; the decision to take this approach was based on the reliability of the data available.

Gender data was collected from either personal knowledge or from the most probable gender being estimated from an internet search. This highlights a flaw of this work, with an assumption of gender being binary and able to be detected from how the person presents in person or their internet presence. This approach was taken as we had insufficient information to be able to determine if a person’s gender was neither male nor female, if their gender is different to how they present, or if their gender has changed or is fluid. We acknowledge that this weakens our gender analysis.

In our analysis, we found that only 21 percent of writers were females, with the remainder being males. A similar ratio was also observed in a report by the computational creativity task force, in which less than one-third of Senior Program Committee (PC) and PC members were female (Cunha et al. 2020). Possible reasons could be that it is more difficult for women and non-binary/gender-fluid people to advance in their careers, perhaps due to external barriers and fewer mentors/role models of the same gender. We note (binary) gender diversity of ICCC keynote speakers.

The impact of such an imbalance would mean that datasets, algorithms and creative systems will pick up on gender biases and perpetuate those further. Even with growing awareness of gender imbalances in STEM, this is a glaring reminder of the need to consider this situation.

Geographical and cultural considerations

There can be cultural implications from the timing of deadlines and other important dates in calls for paper. Table 2 gives important dates from calls for papers, in terms of requirements from authors (submissions of review / camera-ready versions of papers). Some of these dates coincide with significant cultural dates, including:¹

- Chinese New Year (in 2023, Jan 22)
- Lent / Easter (in 2023, February 22 - April 9)
- Holi (in 2023, March 8)
- Ramadan (in 2023, March 23 to April 20)
- Passover (in 2023, April 5 to April 13)
- Cinco de Mayo (May 5)
- Buddha’s Birthday (May 26)
- Juneteenth (June 19)
- The Hajj (in 2023, June 26 to July 1)
- Rosh Hashanah (in 2023, Sept 15 - Sept 17)
- Autumnal Equinox (in 2023, Sept 23)

Deadlines are important for conference organisation. However we can recognise that some people may find it more difficult to submit to, or attend ICCC, based on the above date-based observations. While conference attendees

¹We note that many of these dates change from year to year, and give 2023 dates as a guide, sourced from <https://www.diversityresources.com/>.

Year	Main deadlines	conference dates
2017	March 3 / May 5	June 19-23
2018	March 2 / May 12 / May 25	June 25-29
2019	Feb 28 / April 19 / May 3 / May 6 / May 17	June 17-21
2020	March 8 / May 25	June 29 – July 3 Sept 7–11
2021	April 2 / June 21 / Aug 19	Sept 14-18
2022	Feb 18 / Feb 25 / April 22 / May 13	June 27-July 1

Table 2: Dates in ICCC’17-22 CFP (calls for papers)

would be expected to work around dates that are problematic, and ICCC dates are typically released in advance, in reality the potential date clashes highlighted above can add an additional barrier to participation in ICCC for some people based on their culture or religion.

It is difficult to reliably capture data on nationality(ies) of authors; however we can make objective study of the countries in which people are based when they publish at ICCC, using paper authorship metadata. Table 1 shows the number of distinct countries represented per year in ICCC authorship. Figure 1 groups this data by continent, and records the location of the conference each year. ICCC conference location typically alternates between Europe and non-Europe.

European-based conferences tend to have more European countries represented in author locations, and conferences on the North American continent tend to have higher representation of countries from Asia, Australia and South America. The data are similar to those observed by (Cunha et al. 2020). Small data hinders observations, and comparisons of absolute numbers per continent are misleading; for example, USA is a considerably higher-represented country in the proceedings than Spain, yet Europe counts are higher than for North America.

In Table 1, the number of authors is highest for 2020, corresponding to the highest number of papers (and a mean of

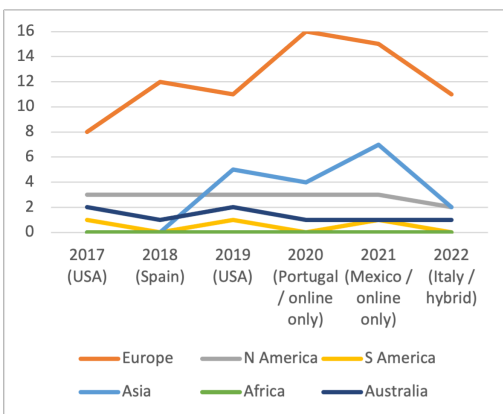


Figure 1: Countries represented in the ICCC proceedings authors per year 2017-22, grouped by continent.

Year	10 Most Frequently Used Phrases
2017	co creative, co creativity, self awareness, creative systems, creative system, creative process, blend space, input spaces, Artificial Intelligence, neural network
2018	creative systems, co creative, game design, story generation, Artificial Intelligence, knowledge base, creative system, aesthetic goal, creative process, meta level
2019	latent space, training data, machine learning, neural networks, co creative, search space, creative systems, knowledge base, design search, Artificial Intelligence
2020	co creative, creative systems, style transfer, co creativity, creative process, human computer, creative system, CC systems, co creation, machine learning
2021	co creative, creative systems, creative process, co creativity, deep learning, language model, punch line, design patterns, machine learning, Artificial Intelligence
2022	GPT 3, fine tuning, C2C-VAE, co creative, natural language, language models, problem solving, Artificial Intelligence, creative systems, creative process

Table 3: Top 10 phrases used over the years in ICCC proceedings (sorted according to frequency - highest to lowest)

2.34 authors per paper, relatively low in an ICCC’17-23 context). 2021 was the only conference to be advertised and held fully online (2020 was only moved online after paper submissions had closed), and the year with highest representation of distinct countries and continents (joint with 2019).

Trends in paper contents

When papers are submitted to a conference for review, it is possible for biases to influence the judgement of reviewers on whether a paper should be accepted or not. While we did not extensively analyze these biases, we did investigate the most frequently used phrases in accepted papers over the years, as shown in Table 3 to gauge if there were any particular trends. We found that certain phrases such as “co-creative”, “creative systems”, and “creative process” were common throughout all years, but in the last two years newer AI technologies were used more in papers. This could be seen in the use of phrases such as “GPT 3” and “deep learning”, suggesting that reviewers may be more inclined to give positive reviews for papers that incorporate newer technologies. However, this could also create more pressure for AI researchers in academia who may not have the resources or means to stay up to date with the latest developments in the field (Togelius and Yannakakis 2023). It could ultimately also lead to the exclusion of researchers.

Effects of online conferences during lockdowns

In 2020 and 2021, ICCC was held online for COVID-19 pandemic reasons. In 2022, the conference was run as a hybrid in-person/online event, with an experience focused on those attending in-person, plus partial online participation options.

With 30 countries represented in author locations in 2021 (the next highest being 24 for 2020), this could represent an emerging trend upwards in the data over time. As 2022 was considerably lower (16 countries), we wait to see whether

the online version of the conference attracted an unusually large representation of authors based in different countries.

An analysis of the unique count of authors over the last 6 years demonstrates that the number of authors and number of papers were also the highest in 2020 and 2021 (see Table 1). A possible reason for this could be that one of the authors of a paper has to register and attend the conference. As conferences were held online in 2020 and 2021, this no longer acted as a barrier to researchers submitting papers. Therefore, the financial cost or the requirement of a visa (if needed) could possibly act as a barrier to paper submission.

Another observation can be made based on the dates in Table 2. Given that the majority of ICCC participants are based in academia, the dates favour those with freedom to travel away from their workplace in June or July. The exception was in 2020 and 2021, where the conferences were moved to September and held as online events.

English as a lingua franca

As is common in academia, for ICCC the *lingua franca*, or mutually adopted language for communication across different nationalities, is English. For non-native speakers of English, it is well recognised that this poses additional difficulty in engaging with academic conferences (Horn 2017).

This is a difficult area to investigate objectively due to lack of access to data and the large scope of such an investigation; however we conjecture that a paper with language issues, such that might arise if someone is writing in a language that they are not fluent in, may be reviewed more critically. This may be due to subconscious bias of reviewers, or even a conscious bias that good command of English language is necessary for ICCC publications.

ICCC'18 gives the only instance of ICCC formally incorporating any language other than English as the main operating language for communication. Held in Salamanca, Spain, the scientific programme includes one workshop (Digital Humanities and Computational Creativity) which was run as a bilingual event, Spanish and English. Papers and presentations were permitted in either Spanish or English, and in the scientific programme² this workshop gives author lists in Spanish (using the Spanish *y* instead of the English *and* to concatenate author lists). This workshop had good engagement, including 6 papers (5 in English, 1 in Spanish) and a round-table discussion. It is worth recalling that this workshop was organised by a group based in Salamanca (the host location). Sadly neither this workshop nor inclusion of multilingual participation have occurred in subsequent years.

Outside of ICCC proceedings, it is worth highlighting the volume on Computational Creativity published in Spanish (Perez y Perez 2015). This book contains an edited collection of Spanish-language papers from frequent contributors to CC research. Sadly this book has only been cited 11 times since 2015 (Google Scholar, last accessed May 2023).

²See <https://computationalcreativity.net/iccc2018/scientific-programme>.

Discussion

The preceding sections highlight the growth of computational creativity as a field of artificial intelligence. Despite its expansion, the diversity of the field has not kept pace.

In the papers we reviewed that discussed creative systems, we found only two instances where measures were taken to mitigate potential biases (Branch, Mirowski, and Mathewson 2021; Khalifa, Barros, and Togelius 2017) and one where the inclusion of cultural bias in the dataset was acknowledged (Mirowski et al. 2022).

In our analysis, we have used only the publication data made publicly available via the proceedings. We acknowledge that this is a limitation of the research; we cannot learn from submissions that were rejected or were not published in the main proceedings, which could be immensely valuable.

Despite the existence of biases and lack of diversity, efforts have been made to make the conferences accessible to a wider audience by alternating the location of the conferences to different continents, as well as the hybrid online/remote format for attendance for ICCC'21.

Conclusions and recommendations

Recommendations emerging from these investigations are:

- Know your data! Collect demographic data?
- In order to promote diversity in computational creativity, it is essential to create datasets that are representative of art from different regions of the world. One such effort is the creation of a dataset of Ukiyo-e, an important style of pre-modern Japanese art (Tian et al. 2021).
- Evaluation is a critical aspect of assessing the effectiveness of creative systems. As such, we recommend that a framework be established for evaluating diversity and bias in creative systems, similar to those already in place for assessing the creativity of a system. This will help ensure that future work in computational creativity is more inclusive and that biases are identified and addressed.
- Steps should be taken to ensure effective guidance and role model representation for women, non-binary or gender-fluid people in earlier career stages, e.g. via the doctoral consortium.
- Some potential authors may face barriers to submitting publications due to language barriers or to clashes between conference dates and religious or cultural dates.
- The hybrid model showed that while the number of unique authors increased, for those attending the conference online the experience could have been better in terms of interaction with peers and senior researchers. Efforts should be taken to ensure that they have similar opportunities as participants attending in-person.
- Our study focuses on *accepted* papers. If access could be arranged for relevant data, a broader follow-up study could provide significant value by including for comparison those papers which were not accepted, or which were accepted for parts of the conference such as the doctoral consortium or demo sessions, but were not consequently included in the proceedings or which were withdrawn.

While CC is growing rapidly, it is becoming increasingly important to ensure steps are taken to increase diversity. We hope that this paper contributes to the growing debate.

Author Contributions

The paper comes from the original ideas of Author 1 (MB), who led the work. These ideas were further developed and expanded by Author 2 (AJ). Both authors contributed to analysis, writing of this manuscript and recommendations.

Acknowledgments

We acknowledge the constructive and thoughtful comments from two anonymous reviewers.

References

- Ackerman, M.; Morgan, J.; and Cassion, C. 2018. Co-creative conceptual art. In *ICCC'18*.
- Banar, B., and Colton, S. Connecting audio and graphic score using self-supervised representation learning—a case study with György Ligeti’s Artikulation. In *ICCC'22*.
- Branch, B.; Mirowski, P.; and Mathewson, K. 2021. Collaborative storytelling with human actors and ai narrators. In *ICCC'21*.
- Burgdorf, K.; Rostamzadeh, N.; Srinivasan, R.; and Lena, J. 2022. Looking at creative ML blindspots with a sociological lens. *CoRR* abs/2205.13683.
- Cunha, J. M.; Harmon, S.; Guckelsberger, C.; Kantosalo, A.; Bodily, P. M.; and Grace, K. 2020. Understanding and strengthening the computational creativity community: A report from the computational creativity task force. In *ICCC'20*.
- Gordon, S.; Mahari, R.; Mishra, M.; and Epstein, Z. 2022. Co-creation and ownership for AI radio. In *ICCC'22*.
- Harris, M.; Harris, H.; and Bodily, P. 2020. ERwEM: Events represented with emotive music using topic-filtered tweets. In *ICCC'20*.
- Hillier, W. 2022. 10 great places to find free datasets for your next project. <https://careerfoundry.com/en/blog/data-analytics/where-to-find-free-datasets/>.
- Horn, S. A. 2017. Non-english nativeness as stigma in academic settings. *Academy of Management Learning & Education* 16(4):579–602.
- Hämäläinen, M.; Alnajjar, K.; and Poibeau, T. 2022. Modern French poetry generation with RoBERTa and GPT-2. In *ICCC'20*.
- Khalifa, A.; Barros, G. A. B.; and Togelius, J. 2017. Deeptingle. In *ICCC'17*.
- Lauzen, M. M. 2020. It’s a man’s (celluloid) world: Portrayals of female characters in the top grossing films of 2019. http://womenintvfilm.sdsu.ex6.https.443.ipv6.faw.com.cn/wp-content/uploads/2020/01/2019-Its_a_Mans_Celluloid_World_Report_REV.pdf.
- Loi, M.; Viganò, E.; and van der Plas, L. 2020. The societal and ethical relevance of computational creativity. In *ICCC'20*.
- Loughran, R. 2022. Bias and creativity. In *ICCC'22*.
- Melistas, T.; Siglidis, Y.; Kalogiannis, F.; and Manouach, I. 2021. A deep learning pipeline for the synthesis of graphic novels. In *ICCC'21*.
- Mendes, R., and Oliveira, H. G. 2020. Comparing different methods for assigning portuguese proverbs to news headlines. In *ICCC'20*.
- Mirowski, P.; Banarse, D.; Malinowski, M.; Osindero, S.; and Fernand, C. 2022. Clip-clop: Clip-guided collage and photomontage. In *ICCC'22*.
- Murai, H.; Toyosawa, S.; Shiratori, T.; Yoshida, T.; Nakamura, S.; Saito, Y.; Ishikawa, K.; Nemoto, S.; Iwasaki, J.; Ohta, S.; Ohba, A.; and Fukumoto, T. 2022. Extraction of typical story plot patterns from genres within Japanese popular entertainment works. In *Proceedings of the 13th International Conference on Computational Creativity, ICCV'22*.
- Navarro, M., and Oliveira, H. G. 2018. Integrating ethno-music and tra-la-lyrics for composing spanish popular songs. In *ICCC'18*.
- Perez y Perez, R., ed. 2015. *Creatividad Computacional*. México D. F: UAM-Cuajimalpa-Patria. https://play.google.com/store/books/details?id=P8_HCgAAQBAJ.
- Stahl, G. K.; Maznevski, M. L.; Voigt, A.; and Jonsen, K. 2010. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies* 41:690–709.
- Tian, Y.; Suzuki, C.; Clanuwat, T.; Bober-Irizar, M.; Lamb, A.; and Kitamoto, A. 2020. Kaokore: A pre-modern Japanese art facial expression dataset. In *ICCC'20*.
- Tian, Y.; Clanuwat, T.; Suzuki, C.; and Kitamoto, A. 2021. Ukiyo-e analysis and creativity with attribute and geometry annotation. In *ICCC'21*.
- Togelius, J., and Yannakakis, G. N. 2023. Choose your weapon: Survival strategies for depressed AI academics. <http://arxiv.org/abs/2304.06035>.
- Walters, T. 2018. A tripartite approach to accessibility, diversity, and inclusion in academic conferences. In *Accessibility, inclusion, and diversity in critical event studies*. Routledge. 230–241.
- wikiart.com. 2023. Artists by nationality. <https://www.wikiart.org/en/artists-by-nation>.