AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

ROADMAP ON AI TECHNOLOGIES & APPLICATIONS FOR THE MEDIA INDUSTRY

SECTION: "AI FOR PUBLISHING: FROM CUSTOMER INTELLIGENCE TO PROSUMER INTELLIGENCE"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Authors	**Julià Vicens** (Eurecat - Centre Tecnològic de Catalunya) **Mihnea Tufis** (Eurecat - Centre Tecnològic de Catalunya) **Cristian Consonni** (Eurecat - Centre Tecnològic de Catalunya) **Patricia Castillo** (Eurecat - Centre Tecnològic de Catalunya) **David Laniado** (Eurecat - Centre Tecnològic de Catalunya) *The authors of this contribution are part of the **Möbius H2020 research project**[1] , which aims to modernise the European book publishing industry by remodelling the traditional value chains and business models uncovering the prosumers potential and delivering new enriched media experiences.*

This report is part of the deliverable D2.3 - "*AI technologies and applications in media: State of Play, Foresight, and Research Directions*" of the AI4Media project.

You can site this report as follows:

F. Tsalakanidou et al., Deliverable 2.3 - AI technologies and applications in media: State of play, foresight, and research directions, AI4Media Project (Grant Agreement No 951911), 4 March 2022

This report was supported by European Union´s Horizon 2020 research and innovation programme under grant number 951911 - AI4Media (A European Excellence Centre for Media, Society and Democracy).

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf.

Copyright

© Copyright 2022 AI4Media Consortium

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the AI4Media Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.
All rights reserved.

[1] Möbius project (funded by H2020 under grant agreement no 957185): https://mobius-project.eu/

AI for publishing: from customer intelligence to prosumer intelligence

Current status

Vendors of products from very diverse sectors have been implementing increasingly sophisticated processes for capturing and analysing data from their customers in order to understand their specific needs, to enrich the catalogue of products, and to iteratively improve the individual relationship with each customer. These processes, denoted customer intelligence, have been favoured precisely by the digitisation of products and interactions with customers. Nevertheless, the consumer is considered as a passive entity, without accounting for their creative potential. The traditional frontiers between vendors and customers in media markets are blurring. Therefore, the times call for a better understanding of the prosumer phenomenon as a key aspect for business competitiveness.

Although user-centric innovation began to be discussed in the 1970's, it is not until very recently that evidence on the impact of involving users in innovation processes for improving business strategies started to arise[2]. The concept of "prosumer" is not new, either. A decade later, in 1980, Alvin Toffler coined the term to signify the people who produce many of their own goods and services, and will tend to replace consumers in the transition from the Industrial to the post-industrial Age.[3] Izvercian et al (2013)[4] define prosumers as usual consumers "*eager to engage, malleable to be transformed and creative enough to produce valuable outcomes*". Importantly, they "*become empowered and are aware of their worth*". Thus, they have potential for collaboration and co-creation of value that may be useful for corporate interests across various industries and markets.

Wattpad online social reading platform is already capitalising on its overwhelming amount of original stories created by their users, and the data they generate in the process. For Wattpad, an open community of writers and readers, prosumers become sources of fresh ideas.[5] The successful Netflix original film "The Kissing Booth" is based on a 2011 story published in Wattpad by a 15-year old that was read by 19 million people on Wattpad before it was turned into a series of books. To keep on this track, Wattpad is partnering with media networks and producers to adapt Wattpad stories for television and has recently launched a Book Division that levers on their Story DNA machine learning technologies for combining data driven approaches "with human editors' critical eyes".[6]

---

[2] Gamble, J. R., McAdam, R., & Brennan, M. (2019). How User-centric Innovation is Affecting Stakeholder Marketing Strategies: Exploratory Findings from the Music Industry. *European Management Review*. https://pure.ulster.ac.uk/ws/portalfiles/portal/12590625/Gamble_et_al._2018_.pdf

[3] Toffler, A., & Alvin, T. (1980). The third wave (Vol. 484). New York: Bantam books.

[4] Izvercian, M., Seran, S. A., & Buciuman, C. F. (2013). Transforming usual consumers into prosumers with the help of intellectual capital collaboration for innovation. *International Journal of Information and Education Technology*, 3(3), 388. http://www.ijiet.org/papers/304-N00018.pdf

[5] C. Lee, How Wattpad Is Rewriting the Rules of Hollywood (2018): https://www.vulture.com/2018/07/how-wattpad-is-rewriting-the-rules-of-hollywood.html

[6] C. de León, Wattpad, the Storytelling App, Will Launch a Publishing Division (2019): https://www.nytimes.com/2019/01/24/books/wattpad-books-publishing-division.html

Another iconic example illustrating the power of prosumer communities is the publishing phenomenon "Fifty Shades of Grey" by author E.L. James. The trilogy was initially a fanfiction creation based on the successful Twilight series (a vampire-themed fantasy romance). After some reworking and interactions with fellow fanfiction followers, James self-published – in e-book and print-on-demand formats – the first installment of her trilogy in 2011 through Australian virtual publisher "The Writer's Coffee Shop." Viral marketing did the rest and "Fifty Shades of Grey" managed to sell 100 million copies worldwide and had its first movie premiered in 2015, only four years after the book was first published.[7]

While current practices of publishers are still mostly based on a vision of the consumer as a passive actor who will just buy or not buy the product, and more or less traditional marketing and recommendation approaches are followed in order to maximise sales, the potential of the prosumers as co-creators of content and in all the steps of the process is not fully taken into account, and the enormous wealth of content and interactions generated by these online communities remains still untapped (Figure 1).

Furthermore, popular fanfiction platforms like Archive of Our Own (AO3) or fanfiction.net, where thousands of users co-create content every day, are based on relatively old technology and offer a simple interface that does not take advantage of the advances in UX research from recent years, or on automated algorithms even for established tasks such as filtering and recommending content to the users. There is room for improving user experience and collaboration practices in this field, with AI at the service of the users and the communities.

In the Möbius project[1], focused on "The power of prosumers in publishing," funded by the European Commission under the Horizon 2020 programme, the aim is to develop methods and tools to effectively streamline cooperation with prosumers – including from open communities – in publishing workflows, and in particular: to leverage on authors and influencers for informing content innovation and publishing decisions and establish cooperation strategies; and on beta-testers and early-adopters for innovation processes for new products or experiences.

Figure 1: In fanfiction communities, prosumers build on existing work of fiction to write new stories in a canonical fictional universe, such as the one from Harry Potter, represented in the figure. Photo by Jules Marvin Eguilos on Unsplash[8].

[7] Statista, Number of copies E.L. James's 'Fifty Shades of Grey' sold in selected countries worldwide as of February 2014: https://www.statista.com/statistics/299137/fifty-shades-of-grey-number-of-copies-sold/
[8] Image source: Jules Marvin Eguilos, Unsplash, https://unsplash.com/photos/YrcfSDVli3Y

Research challenges

In the publishing sector, taking advantage of the wealth of data created by prosumers, in the form of original content, reviews, feedback and interactions, implies being able to mine this data, make sense of it and extract actionable knowledge for improving the community dynamics, productivity and synergies, and for ultimately publishing and promoting successful and innovative products.

While a big effort has already been spent in previous research to develop methods and algorithms for mining social media and online platforms, some particularities of the context of fanfiction and prosumer communities make it unique, so that it needs to be addressed by a specific modelling effort.

This includes identifying and characterising different prosumer profiles, according to the capacity in which they are making a contribution. As a first classification, we can see:

- *Authors* create new stories, even if they are fandom contents.
- *Influencers* may also be generating original content, although not to create new stories, but to generate discussion about them; and so they are able to engage communities and elicit contributions around this shared interest.
- *Beta-testers* take part in the development of new products and services, their feedback is crucial for ensuring the innovation meets user expectations, and that it is meaningful, useful and acceptable.
- *Early-adopters* are the groups most likely to adopt or consume an innovation as soon as it is released, so their reaction to the near-to-final product is of utmost interest. Importantly, they all bring user knowledge, interactions, opinions, and help generate further user requirements that would typically elicit new user-driven innovation processes.

Identifying different profiles, roles, and levels of involvement and reputation of prosumers is crucial for being able to engage specific users for specific goals and tasks. Furthermore, it is part of mining community dynamics in order to monitor the state and health of the community, and to be able to take action to improve it.

Collaboration and co-creation happen spontaneously in prosumer communities, however the interaction mechanisms allowed by the platform play an important role: they may foster more or less fruitful cooperation dynamics, encourage synergies and distributed mentorship,[9] or in some cases on the contrary they may even hinder collaboration, or incentivise toxic behaviour. AI in this sense can help experiment and find out what works better, not only with empirical analysis of digital traces from existing communities, but also with practices such as A/B testing under tight ethical codes.[10]

[9] Campbell, J., Aragon, C., Davis, K., Evans, S., Evans, A., & Randall, D. (2016). Thousands of positive reviews: Distributed mentoring in online fan communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 691-704).

[10] Benbunan-Fich, R. (2017). The ethics of online research with unsuspecting users: From A/B testing to C/D experimentation. *Research Ethics*, 13(3-4), 200-218.

Knowledge extracted from the interactions and community dynamics can be complemented by mining the textual content of the works produced by the users, extracting from the text features such as topics, emotions, characters and entities, and linguistic styles, quality, and diversity. Combining textual features extracted from the content with interaction features and social dynamics around the corresponding content may open up a great potential for innovation, where AI can help make the co-creation process more effective and satisfactory for prosumers.

We believe in this context that not only a human-centric approach[11] is fundamental, ensuring algorithmic transparency and explainability, and keeping a watchful view on bias and discrimination potentially introduced by the systems, so that AI is actually at the service of the user, who is able to make sense of the algorithmic decisions; but also a community-centric approach, "keeping community in the loop" as proposed in recent research on the Wikipedia community, to make sure algorithms are aligned with the values and needs of the community.[12]

Societal and media industry drivers

Vignette: Publishing a manuscript from an online fanfiction community

Ginette works at a publishing company, and her duty is to select manuscripts for publication. Ginette does not only examine manuscripts she receives, but she works in collaboration with the online community around a co-creation platform, and relies on an AI monitoring system that discovers promising content.

Among the hundreds of works that are created every day on the platform by the users, the system points Ginette to a manuscript that has been created just a few hours earlier, and has already received positive feedback from an influential trend-setter from the community.

Based on the very first feedback, and on the linguistic, topical and emotional features automatically extracted from mining the manuscript's text, the system has selected this as a promising manuscript, and shows Ginette a bunch of synthetic indicators estimating its linguistic style, quality, emotional content, main topics, main characters from the original book from which it is inspired, relation to other manuscripts published by the company, potential target audience in terms of demographic characteristics and other marketing criteria.

With a look at the indicators, Ginette immediately sees an opportunity in this manuscript, and activates the "early feedback" program. The system identifies the users who have the best match with the features of the story, and at the same time have a relevant experience and position in the network of users of the platform, and invites them to review the manuscript. So, beyond the spontaneous reviewing process that will happen within the community, these users are pro-actively invited to give their feedback. The users will have the possibility to give suggestions to the author, and eventually become co-authors.

[11] Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *interactions*, *26*(4), 42-46.

[12] Smith, C. E., Yu, B., Srivastava, A., Halfaker, A., Terveen, L., & Zhu, H. (2020, April). Keeping community in the loop: Understanding wikipedia stakeholder values for machine learning-based systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Once various rounds of reviews are completed, the book enters the final phase of the preparation for publication by the publishing company. The book that is launched to the market is in a way a product of the whole community that is involved also in creating images and messages for promoting it in social media.

Future trends for the media sector

AI may help to develop prosumer intelligence for the publishing sector, by providing algorithms and techniques for taking advantage of the untapped potential in the wealth of data generated by the users, and leveraging it for improving the co-creation and dissemination processes. This can be achieved addressing several issues and goals:

- *Early detection of trends in content consumption*: Identify content and topics that will become popular in the near future, and that have potential for publishing; with algorithms based on temporal patterns, interactions, and features of user generated content.
- *Early detection of trends in content production*: identify topics, styles, emotions, genres and subgenres, patterns that are increasingly adopted by authors, with particular attention to users that tend to establish or adopt early new trends that will spread in the community.
- *Fostering synergies and distributed mentoring*: identify users that could fruitfully collaborate with each other or review one another's work, based on common interests, and complementary abilities, expertise and styles.
- *Identification of relevant users*: mine user behaviour to identify prosumers that assume specific roles in the community: authors, fans, influencers, beta-testers, early-adopters, trend-setters.
- *Selection of seed content*: identify content and stories, from the community itself as well as from external sources, which can be used as input to stimulate the co-production of new content.
- *Monitoring of community dynamics*: identify phases and trends in the growth of a prosumer community, and actions needed to improve community health.
- *Annotation mining*: enrich the reader's experience not only based on feedback, comments and annotations from other users, but also on further predictions based on such user generated content.
- *Indexing of content*: automatic identification and indexing of topics, characters, objects and other kinds of entities, in order to allow for browsing and exploring content, and for connecting works and manuscripts with each other at different levels of granularity, including by chapter or paragraph.

Goals for next 10 or 20 years

When we consider prosumer communities devoted to co-creation of content, such as fanfiction communities, we have to be aware that at the very core of these socio-technical systems we have humans, both at the individual level with their creativity beyond their personal tastes, preferences and habits, and at the community level with their interactions, their social dynamics, their collaboration practices, and their belonging to a social group with shared values

and objectives. Therefore, we believe the most crucial goals for AI in this field have to do with being able to put algorithms at the service of users and communities, integrating into their daily practices and aligning with their needs and values. AI algorithms can act as a multiplier of inequality and reinforce existing bias, therefore they must be designed keeping in mind their effect on the people who use them or who may be affected, which must be able to participate and control how the AI is used. We envision a shift from currently dominant black-box models and opaque algorithms run by the platforms to maximise variables such as clicks or time spent on the platform, to an emerging paradigm of human-centric and community-centric AI at the service of the communities, with transparent algorithms that are able to foster social connections, synergies and human creativity.

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu