Immersive Automation: summary of project outcomes

The Immersive Automation-project is a research consortium where data scientists, linguists, and journalists worked together with media companies to bring forward a solution for producing engaging and data driven content. Our aim was to create a roadmap and demonstrate a future news ecosystem based on automated storytelling, audience engagement, and user experience. The duration of the project was December 2016–May 2018.

First, we present a short summary of the main parts of the project.

1. Data specification: Identification and categorisation of data sources. Special focus on identification of relevant structured data for editorial use.

2. Automation of news production: Development of methods for automated text generation, as well as provision of models for the structure of news stories.

3. User engagement: User tests for the automatically generated content.

4. News ecosystem: Development of an ecosystem where technical, organizational, as well as commercial levels are both conceptualised and demonstrated.

5. Development of skills: Training selected journalists in computational thinking, algorithmic news production, and editorial tools based on automation.

Main achievements in the project:

1. **Technical development (Valtteri 1.0 and 2.0)**

During the project, we produced an abstract architecture for largely domain- and language independent natural language generation. This architecture was then first implemented in early 2017 as a tri-lingual system producing news about the 2017 Finnish municipal elections. The system was presented to the public the night of the election and garnered interest from both users and media.

During late 2017 and early 2018, the system was transferred to the significantly more complex domain of crime statistics, producing news on that topic in two languages. In addition, the system received improvements to its language production and data analysis capabilities. This second implementation also enhances the produced news articles with visual components, such as automatically generated interactive graphs and maps.

The two systems demonstrate the strengths of automated news generation by analysing and evaluating very large datasets, and then by creating hundreds of thousands of articles – both requiring impossible amounts of resources if conducted by hand.

During late 2017, additional trials were made for increasing the multilinguality of the system by taking advantage of existing machine translation systems. It was determined, however, that the current state-of-the-art machine translation is not of sufficiently high quality to employ the trialled methods, especially insofar as the Finnish language is concerned.

2. **Acceptance of automatically generated texts**

User tests produced new knowledge related to acceptance of automatically generated news. Particularly the second user test confirmed that the technical language improvements done in Valtteri 2.0 lead to improved article quality. Initially, the audience testing system was specified, and approved by the steering group. After this, the test environment was built, and the steps of the test were defined. A test panel

comprising of 152 users were recruited. The panel was a representative sample of the Finnish population. Users were not informed of which texts were automatically generated and which were human-written. All the articles were evaluated along four dimensions: credibility, liking, quality and representativeness. As expected, the articles written by Valtteri received lower ratings than those written by journalists, but overall the ratings were satisfactory (avg. 2.9 vs. 4.0 for journalists on a 5-point scale). Valtteri's best rating (3.6) was for credibility. The computer-written articles that the evaluators could freely select received slightly better ratings than the preselected computer-written articles. When looking at the results by demographic groups, males aged 55 or more liked the automatic articles best, whereas females aged 34 or less liked them the least. Evaluators mistook 21% of the computer-written articles as written by humans, and 10% of the human-written articles as computer-written. The share of users making these mistakes grew with age. Overall, male evaluators made less writer-identification mistakes than female evaluators did. The results are published in the scientific journal IEEE Access Melin, M., Bäck, A., Södergård, C., Munezero, M. Leppänen, L. & Toivonen, H. (2018): "No landslide for the human journalist – An empirical study of computer-generated election news in Finland. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8424161

After this, we also tested the so-called Crime Valtteri 2.0 with the same panel.

3. Training for journalists

The training sessions were an important part of the project. We aimed at educating journalists in computational thinking and related issues, such as the availability and potential use of open data. We also involved journalists in workshops focusing on creating chatbots for editorial purposes and learning the basics of machine learning, for example. The feedback from the journalists was positive. We observed a conflict between the short-term impact (new tools and techniques learned) and long-term impact (a vision of what radical transformation of news work means in five to ten years).

4. Spreading the message

Dissemination of research findings was successful and we presented our project in a number of conferences and seminars, mostly internationally. Our state-of-the-art report for WAN-IFRA, the global trade organisation for the publishing industry, spreads new knowledge effectively on a global scale. WAN-IFRA has a network of 3,000 news publishing companies and technology entrepreneurs, as well as 80 member publisher associations representing 18,000 publications in 120 countries. All members receive the report.

5. Public visibility and impact

We also observed that the publicity around the Immersive Automation project in Finnish media and industry publications ignited Innovation around news automation in the Finnish media sector. When Valtteri was launched in April 2017, two news organisations – Helsingin Sanomat and the Finnish Broadcasting Company Yle – also presented their own services. There has been commercial interest both in Finland and abroad to implement the software that drives Valtteri in other domains and countries. The team was involved in serious negotiations to commercialize Valtteri, but decided to withdraw when funding for another three more years had been secured. Thus, the plans to develop a commercial version of Valtteri never materialised.

6. The research continues

One measure of success is the capacity to gain funding. The two teams from the University of Helsinki have continued to work closely together and are now part of a large European consortium, Embeddia, which

received 3 million euros in Horizon 2020 funding in August 2018. Two other applications have gone to the second round in calls for funding.