ICT with Industry 2024
15-19 January 2024, Lorentz Center @Oort

Under the umbrella of ICT-research Platform Netherlands (IPN) the workshop ICT with Industry 2024 was organized by NWO in collaboration with the research schools ASCI, IPA and SIKS.

Scientific
The workshop strives for direct and rapid interaction between ICT researchers and industrial partners with the following objectives:
➔ To stimulate contact between ICT research and industrial R&D.
➔ To obtain creative solutions for industrial problems and to find new approaches that could lead to such solutions.
➔ To give insight into the wide range of possibilities ICT research offers and thereby enable accelerated innovation.
➔ To enrich the PhD students' and postdocs' experience in collaborating with industry.

The ICT with Industry 2024 workshop took place between January 15th and 19th 2024. A total of 25 participants from industry and academia were working on these cases, of which about 40% female participants. The cases for this edition had no specific topic and all of them resulted in promising methods and concrete directions to further develop. For future collaborations NWO provides several funding instruments, e.g. Perspectief, Take Off, etc. The teams were invited to present during the NWO showcase track at ICT.OPEN2024.

Outcome of cases
Axini: AI/LLMs and Model Based Testing
Participants: Shengyuan Yan, Vasileios Tsouvalas, Felix Frohnert, Arina Kudriavtseva, Dolly Sapra (academic lead) and Machiel van der Bijl (case owner)

The use-case with Axini in the ICT with Industry 2024 workshop focused on exploring the application of AI and Large Language Models (LLMs) in model-based testing. Participants engaged in practical exercises and discussions, centering on the Axini platform and its conceptual framework for model-based testing. The workshop aimed at developing models using LLMs, specifically targeting the creation of models using the Axini's propriety modeling language (AML).

Key activities included prompt engineering, rule setting, model creation, and verification. Initial attempts showed varying success rates across different versions of GPT, indicating the evolving capabilities of LLMs in understanding and generating syntactically correct models. The workshop progressed through collaborative efforts involving the uploading of various models and specifications, leading to the identification and correction of errors by the LLMs.

Significant insights were gained in the process of converting natural language specifications into AML standards. The workshop highlighted the importance of iterative learning and progressive corrections to teach the GPT about the nuances of accurate AML script. The participants successfully utilized a combination of Python and Axini's platform to generate model visuals, gradually expanding the model's functionalities and aligning them with specified requirements.

The workshop concluded with a set of observations and recommendations for future model-based testing using LLMs. It underscored the importance of detailed comments in code for better solution alignment and the effectiveness of teaching syntax and logic to improve LLM responses.

BrainCreators: Exploring efficient interactions of human- and machine intelligence for infrastructure inspection: Road crack segmentation with deep learning
Participants: Kirsten Maas, Sanne van der Linden, Gizem Karagoz, Pankaj Kumar, Sjoukje Osinga (academic lead), Dilli Paudel (academic lead) and Maarten Stol (case owner)

Automated recognition of road distress (cracks), is of growing importance in infrastructure maintenance. Pixel segmentation can be used for severity estimation of road cracks, but obtaining ground truth segmentation masks is labour intensive. As an alternative, foundation models like Segment Anything Model (SAM) may be finetuned using small datasets only (few-shot learning). Prompt engineering replaces labelling strategies. Central question is: How efficient is prompting in terms of achieved annotation quality?
We explored SAM by establishing an 'uninformed' baseline through random sampling: class balanced (prompts were pixels selected uniformly from positive (part of a road crack) and negative (not part of a crack) classes). We then pursued various 'beyond uninformed' prompting strategies to improve the baseline's results: manual (including different order); negative area selection; different road types; branching selection; edges selection; mislabelled regions selection (Sofiiuk 2022) and using an augmented SAM model. For data, we used the open source dataset crack500, which contains a variation of road types and crack shapes, and well-defined masks with varying widths (no 'pixelization'). We picked 5 representative cases. We use IoU (Intersection over union) as a metric to compare the pixel masks with ground truth.

Result showed that our experiments did not bring ground breaking improvements on IoU scores. Our main conclusion is that SAM is not the optimal tool to identify crack segments, with a similar level of performance as for objects and people. Future research may include to change contrast in images, to finetune SAM specifically on cracks, and to investigate other models.

eScience center: Semantic continuity

Participants: Emanuele Mezzi, Qi Huang, Miltiadis Kofinas, Osman Mutlu, Vidya Prasad, Niki van Stein (academic lead) and Elena Ranguelova (case owner)

In the Lorentz Workshop "ICT with Industry" the eScience team investigated the concept of semantic continuity in explainable AI (XAI), which postulates that semantically similar inputs should yield similar explanations from XAI methods on the AI models (given that the AI model gives similar predictions). The team developed a new metric to measure semantic continuity and explored its application in image and text domains through experimental setups involving datasets of shapes and text with varying semantic scenarios. Results are analyzed using statistical measures to understand the relationship between semantic changes in inputs and the corresponding explanations, aiming to enhance the explainability and reliability of AI systems.
The team succeeded in analysing the XAI method GradCam for an image classification task and RISE for a sentiment analysis task on text data. After the workshop, the team aims to further develop the proposed methods into a full paper to be submitted to the World XAI conference.

Axini: Testing 3D racing game

Participants: Arwin Gansekoele, Xiaotian Guo, Vladyslav Andriiashen, Paris Mavromoustakos Blom, *Michel Chaudron (academic lead) and Machiel van der Bijl (case owner)*

3D videogames are complex software products, and their scope and desired quality is quickly growing in recent years. Testing a videogame is a challenging task that traditionally involves a large number of professional QA experts and feedback from users. However, a continuous development process and tight schedule often lead to unsatisfactory testing coverage and game quality. Model-based testing is widely used as a method of automating the testing process. It assumes that the System Under Test (SUT) could be represented as a finite-state machine and different test cases could be formulated as a sequence of transitions between states. We introduce a general approach of connecting a videogame to the model-based testing platform. We propose to formulate SFM states using information extracted from the game engine and discuss how to select information based on the test. The crucial step is the introduction of Controller – a system that translates a high-level in-game behavior into basic control commands. This system is used to describe transitions between states using not only standard videogame controls but more complex primitives. We have demonstrated the proposed approach to a problem of movement testing in an open-source 3D racing game. Our method was able to mitigate the problem of synchronization between the game and the testing platform. The Controller could be used to produce a large variety of complex in-game behavior. We discuss the fundamental issues of formulating a videogame model and the generalizability of our approach.

- Videogame production is a quickly growing industry developing complex software products
- Testing a videogame is a challenging task that usually involves a large number of human testers
- Model-based testing presents an approach to automate and categorize testing processing by representing the System Under Test (SUT) as a finite-state machine
- We have proposed an approach of representing information from the game as states and game inputs as transitions.
- The crucial step is the introduction of the Controller - a system that translates a high-level in-game behavior into basic control commands.
- Focus on high-level commands mitigates synchronization issues between the testing platform and the SUT.
- Our approach was demonstrated using a 3D racing game as an example.

- The in-game movement was tested using collision status as a state with the Controller capable of in-game steering and pathfinding.