

# HAVA-Lab: Human-Aligned Video AI Laboratory

הה (hava) - breathing, living

Lead Principal Investigator:	prof. dr. Cees Snoek ( <a href="mailto:c.g.m.snoek@uva.nl">c.g.m.snoek@uva.nl</a> )	FNWI
Lab Manager:	dr. Pascal Mettes ( <a href="mailto:p.s.m.mettes@uva.nl">p.s.m.mettes@uva.nl</a> )	FNWI
Co-Principal Investigators:	dr. Erwin Berkhou (e.berkhou@acta.nl) prof. dr. Tobias Blanke ( <a href="mailto:t.blanke@uva.nl">t.blanke@uva.nl</a> ) dr. Iris Groen ( <a href="mailto:i.i.a.groen@uva.nl">i.i.a.groen@uva.nl</a> ) dr. mr. Heleen Janssen ( <a href="mailto:h.l.janssen@uva.nl">h.l.janssen@uva.nl</a> ) prof. dr. Marie Lindegaard ( <a href="mailto:MRLindegaard@nscr.nl">MRLindegaard@nscr.nl</a> ) dr. Stevan Rudinac ( <a href="mailto:s.rudinac@uva.nl">s.rudinac@uva.nl</a> ) prof. dr. Marlies Schijven ( <a href="mailto:m.p.schijven@amsterdamumc.nl">m.p.schijven@amsterdamumc.nl</a> )	ACTA FGW FNWI FdR FMG FEE AUMC

## Keywords

Video, Artificial Intelligence, Human Alignment

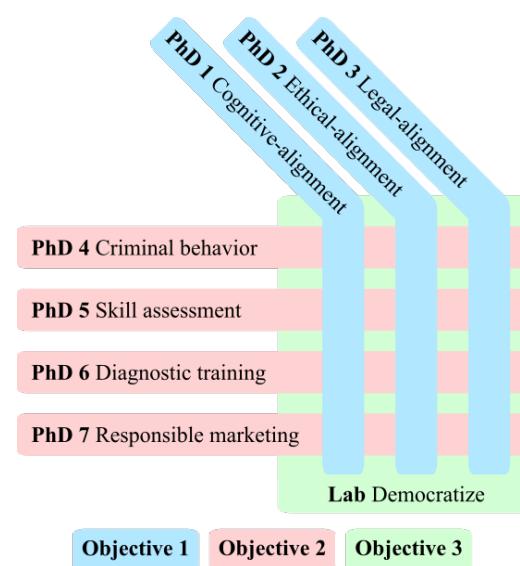
## Research narrative

We propose HAVA-Lab, a research program that equips Artificial Intelligence (AI) for automated interpretation of video data with the ability to align with human values. Video-AI holds the promise to explore what is unreachable, monitor what is imperceivable and to protect what is most valuable. This is no longer wishful thinking. Broad uptake of video-AI for science, for wellbeing, and for business awaits at the horizon, thanks to a decade of phenomenal progress in deep learning [1]. However, the same video-AI is also accountable for self-driving cars crashing into pedestrians, deep fakes making us believe misinformation, and mass-surveillance systems monitoring our behavior. The research community's over-concentration on recognition accuracy has neglected human-alignment for societal acceptance [2]. To enable a much-needed responsible digital transformation, we propose the first of its kind laboratory that studies video-AI from a multi-disciplinary perspective addressing the question:

*What defines human-aligned video-AI, how can it be made computable, and what determines its societal acceptance?*

The state-of-the-art in video-AI is to recognize objects, activities, and their interactions by formulating these recognition problems as deep learning algorithms [3,4]. Simply put, these neural network algorithms learn associations between labels and video-fragments at training time that allow them to predict labels in unseen video at test time. Yet it is becoming increasingly clear that video-AI may perform well in the lab, but is still brittle when deployed under real-world situations where conditions differ from those during training [5]. An effective solution is to simply scale-up the labels and video data seen during training, but this leads to an intellectual and societal dead end. Not only because of the unsustainable compute expenses, but also as labels may inherit intrinsic bias. Especially in Europe, society is unlikely to continue to accept the ethical cost of storing and processing massive amounts of video data without consent of those recorded. Therefore, to make video-AI deliver on its big promise, human-alignment is key.

This proposal enables human-aligned video-AI by addressing three research objectives. Objective 1 will focus on the development of human aligned values in video-AI, while Objective 2 will focus on algorithmic development of video-AI with human alignment for real-world scenarios. Both objectives form a continuous feedback loop; developing new alignment will drive algorithmic development, algorithmic development will shed new light on where alignment is most needed. Objective 3 strives to democratize human-aligned video-AI by creating new knowledge, bringing the knowledge creation efforts at all UvA-Faculties to a higher level, and attracting and developing new talent. The figure shows the connections between the objectives.



#### *Research objective 1: Human alignment in video-AI*

The goal is to incorporate human-alignment during the development lifetime of video-AI algorithms from cognitive, ethical, and legal perspectives, each led by a PhD student.

- *Cognitive-alignment (w/Groen FNWI)*. Can video-AI become more computationally efficient by making it resemble the human brain? The rich perceptual content of video, with objects moving and activities unfolding in space-time, poses a much more challenging cognitive-alignment problem than work focusing on static images only [6]. We expect dynamic adaptation [7] and sparser network geometries [8] to better approximate human efficiency.
- *Ethical-alignment (w/Blanke, FGw)*. How to embed ethical values better in video-AI algorithms? Based on the method of Jaton [9], we will systematically record moments of ‘hesitation’ in the production of video-AI models to understand where genuine choices in the sense of pragmatic morality lie, how to represent these genuine choices in the video-AI algorithms, and how to develop new standards for labeling that include these choices.
- *Legal-alignment (w/Janssen, FdR)*. This PhD project will consider video-AI compliance with fundamental rights and ethical values our European societies are based on. Can we incorporate privacy [10] and legal standards of non-maleficence, equity, or justice by design? Can we develop human-aligned video-AI that accords with legal and regulatory concerns, while grounding legal and policy discussions in technical realities?

#### *Research objective 2: Human-aligned video-AI for UvA*

The goal is to develop video-AI with desired human alignment embedded and usable for the non-expert. We focus on four novel use cases directly relevant for research and education at UvA.

- *Criminal behavior (w/Lindgaard, FMG)*. Video-AI has the potential to detect crime from camera recordings and provide insights into crime-types that are currently unregistered and unreported. However, specialist insights are still required to deal with human biases present in existing data. This requires developing insights into what can be used for detection and new video-AI algorithms recognizing crime without perpetuating unwanted biases.
- *Skill assessment (w/Schijven, AUMC)*. How can video-AI become robust and safeguarded against annotation biases when it comes to real-world deployment, especially in medical settings when it comes to assessing skills? And how should video-AI algorithms be developed to be ethically sound in production when trained under skillslab conditions? We will investigate the step from lab to real world for video-AI constrained by ethical constraints on fairness across all relevant dimensions.

- *Diagnostic training (w/Berkhout, ACTA)*. Diagnostic training is a critical aspect of dental education, as students often struggle with interpreting radiographs. To improve the quality and efficiency of diagnostic training, video-AI can play a crucial role. We will research video-AI algorithms to compare learning strategies and provide insight into diagnostic failures by recognizing the student's behavior and gaze patterns during diagnostic assessment, creating more proficient future dentists.
- *Responsible marketing (w/Rudinac, FEE)*. A central theme in social media analysis is organized online campaigns, both political and economical. In such content, text and image recognition is common, yet videos are often ignored while being key when it comes to virality or synchronized campaigning. We will investigate the role of videos in such social content and the development of video-AI algorithms for detecting when video content has viral potential through domain-specific guidance and alignment.

#### *Research objective 3: Democratize human-aligned video-AI*

Knowledge transfer of human-aligned video-AI happens at the DSC. We will organize a monthly HAVA-Faculty-focus where the research of one student is highlighted together with an invited speaker. Twice a year we organize a HAVA-workshop with tutorials and hands-on notebooks, to share human-aligned video-AI methodologies and practices, and we will organize a yearly 1-day HAVA-conference to share our research and have round table discussions with the broad UvA-community interested in human-aligned video-AI.

#### **HAVA-Lab organization and embedding with DSC**

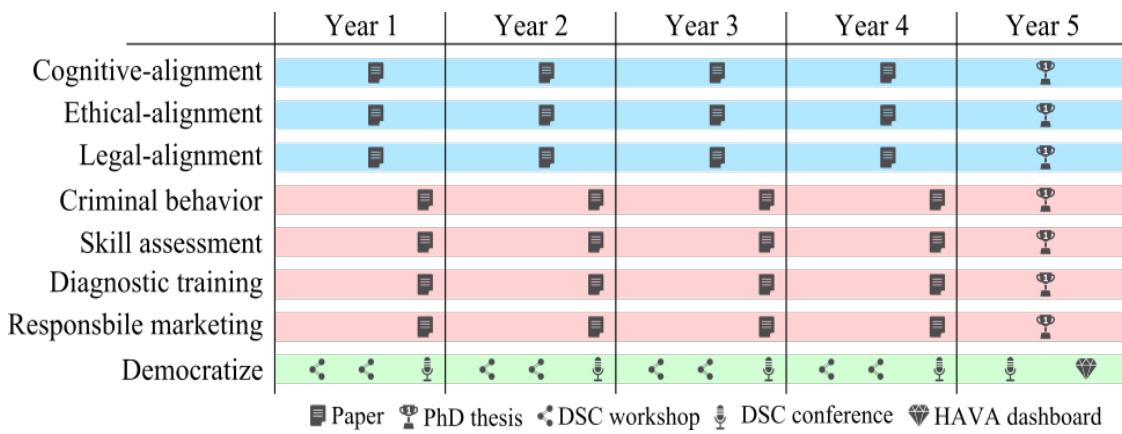
We model the HAVA-Lab along the format of the labs of the Innovation Center of Artificial Intelligence (ICAI), be it with a fresh flavor that accommodates not only technological depth, but especially the full societal breath and responsible use that AI at the University of Amsterdam, and the Amsterdam AI ecosystem in general, has to offer. We consider it crucial that the students sit together in the same collaborative space, so they can inspire each other and get up to speed on general video-AI methodologies quickly, which can be easily transferred from the VIS lab from Snoek, e.g. [11-15]. Hence, we envision that in year one the students will spend 4-days a week in the HAVA-Lab and 1-day a week in the respective Faculty-specific research unit. During the course of their PhDs, we expect the students to increase their visits to the Faculty-specific research unit. Daily organization of the HAVA-lab will be led by our lab manager (tenured co-PI Pascal Mettes).

#### **Interdisciplinarity, diversity and junior faculty**

The HAVA-team is highly multi-disciplinary, as evidenced by our coverage of expertise from all seven UvA Faculties, yet we have found each other in our shared research interest in human-aligned video-AI. The diversity of the team is further evidenced by a balance in gender, nationalities, and career-stages (PD, UD, UHD, HL, UHL), which we deem advantageous for hiring a similarly diverse pool of interdisciplinary PhD students. Junior faculty will have an active role in the individual projects as well as in running the lab. Three of our (junior) team-members have experience with supervising a DSC-PhD project through last year's PhD call. Four of the seven PhD projects build on existing FNWI-collaborations with our Co-PIs from ACTA [16], FEE, and FMG [17] while we have also established new ones with Co-PIs from AUMC, FdR, and FGw.

## Outcomes

The lifetime of the HAVA-Lab is 5 years. PhD students are hired for a 4-year period, but experience teaches us that it will not be possible to let them all start on the same day. We do expect each student to start in year 1. We envision the PhDs to have a socio-technical profile, with the goal to publish in technical-AI venues like CVPR, NeurIPS, multi-disciplinary venues like ACM Multimedia, ACM FAccT, as well as discipline-specific venues relevant for their specific project. Besides publications and PhD theses, we expect to deliver the software and data that are part of each project. Together with the support of the community management and training support from DSC, we will culminate the outcomes of the lab into a virtual dashboard, where key findings are visualized, summarized in blogpost-style, and with technical advances handed out as easy-to-use notebooks. The Gannt chart of the HAVA-Lab is shown below.



To assure feasibility all PhD students will also be embedded at Snoek's VIS Lab at FNWI, with 35 fte working on computer vision and machine learning, as well as the group of their Co-PI at one of the other UvA-Faculties. Collaboration will be stimulated right from the start. For objective 1, addressing all forms of alignment may prove too ambitious. To ensure success, we will continuously evaluate which types of alignment can realistically be made computable at what stage of video-AI algorithm development, be it at pre-training, training and/or deployment. For objective 2, a risk is insufficient availability of video data to train and evaluate the human-aligned video-AI. In that case, we will rely on more videos in existing large-scale collections such as Kinetics, HowTo100M, and Ego4D, and emphasize on few-shot learning and domain generalization tactics for which expertise is available in-house [18,19]. A risk for objective 3 may be that human-aligned video-AI is still too complex for novice data scientists interested in the topic. To assure uptake we will continuously monitor and evaluate usability during our tutorials with hands-on notebooks.

## External matching and spin-off

In-kind matching is provided by each UvA-Faculty in terms of supervision and contribution to the democratization of human-aligned video-AI. Snoek's VIS lab will equip each student with a 2K laptop so they can easily work across locations. We have further assured in-kind access to compute infrastructure in VIS lab (75 1080Ti GPUs, 36 A6000 GPUs, 200TB data storage) and the Informatics Institute (100 1080Ti GPUs). For the yearly workshops and conferences we will request extra funding from ELLIS Unit Amsterdam's support program for events. It is expected that the HAVA-lab research will act as a multiplier for even more intense collaboration across UvA-disciplines, with joint-proposals already planned for the NWO Call on Collaboration between Humans and (semi-)Autonomous systems and NWO's Open Technology programme, as well as personal grants innovating with video-AI outside computer science. We also anticipate new spin-off company's bringing human-aligned video-AI from the lab to society at large.

## References

- [1] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *Nature* 521, 436–444. 2015.
- [2] B. Christian. *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton & Company, 2021.
- [3] J. Carreira, A. Zisserman. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In *Computer Vision and Pattern Recognition*, 2017.
- [4] C. Feichtenhofer, H. Fan, J. Malik, K. He. SlowFast Networks for Video Recognition. In *International Conference on Computer Vision*, 2019.
- [5] FM Thoker, H. Doughty, P. Bagad, CGM Snoek: How Severe is Benchmark-Sensitivity in Video Self-Supervised Learning?. In: *European Conference on Computer Vision*, 2022.
- [6] I. Sucholutsky I & TL Griffiths TL. Alignment with human representations supports robust few-shot learning. *ArXiv*, 2023 <https://doi.org/10.48550/arXiv.2301.11990>
- [7] IIA Groen, G Piantoni, S Montenegro, A Flinker, S Devore, O Devinsky, W Doyle, P Dugan, D Friedman, N Ramsey, N Petridou, JA Winawer. Temporal dynamics of neural responses in human visual cortex. *The Journal of Neuroscience* 42(40):7562-7580, 2022.
- [8] T. Long, P. Mettes, HT Shen, CGM Snoek. Searching for actions on the hyperbole. In *Conference on Computer Vision and Pattern Recognition*, 2020.
- [9] F. Jaton. Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application. *Big Data & Society* 8(1), 2021
- [10] SR Klomp, M. van Rijn, RGJ Wijnhoven, CGM Snoek, PHN de With: Safe Fakes: Evaluating Face Anonymizers for Face Detectors. In: *IEEE International Conference on Automatic Face and Gesture Recognition*, Jodhpur, India, 2021
- [11] H. Doughty, CGM Snoek: How Do You Do It? Fine-Grained Action Understanding with Pseudo-Adverbs. In: *Computer Vision and Pattern Recognition*, 2022.
- [12] Y. Zhang, H. Doughty, L. Shao, CGM Snoek: Audio-Adaptive Activity Recognition Across Video Domains. In: *Computer Vision and Pattern Recognition*, 2022.
- [13] J. Zhao, *et al*: TubeR: Tubelet Transformer for Video Action Detection. In: *Computer Vision and Pattern Recognition*, 2022,
- [14] P. Bagad, M. Tapaswi, CGM Snoek: Test of Time: Instilling Video-Language Models with a Sense of Time. In: *Computer Vision and Pattern Recognition*, 2023.
- [15] FM Thoker, H. Doughty, CGM Snoek: Tubelet-Contrastive Self-Supervision for Video-Efficient Generalization. *ArXiv*, 2023.
- [16] M. van Spengler, E. Berkhout, P. Mettes. Poincaré ResNet. *ArXiv*, 2023.
- [17] W. Bernasco, E. Hoeben, D. Koelma, L. Suonperä Liebst, J. Thomas, J. Appelman, CGM Snoek, M. Rosenkrantz Lindegaard: Promise Into Practice: Application of Computer Vision in Empirical Research on Social Distancing. In: *Sociological Methods and Research*, 2023. *In press*.
- [18] T Kasarla, GJ Burghouts, M van Spengler, E van der Pol, R Cucchiara, P. Mettes. Maximum Class Separation as Inductive Bias in One Matrix. In: *Neural Information Processing Systems*, 2022.
- [19] Z. Xiao, X. Zhen, S. Liao, CGM Snoek: Energy-Based Test Sample Adaptation for Domain Generalization. In: *International Conference on Learning Representations*, 2023.

## Response to the questions by the review committee

*Thank you so much for your UvA Data Science Centre Interdisciplinary Lighthouse Project Proposal [1]. We received 12 excellent proposals in total. Reading them and seeing the teams assembled, I am continually amazed by the creativity, talent, and capabilities of my colleagues here at the UvA. Congratulations! Your proposal was ranked number 1. The committee found your proposal compelling in its boldness of methods, interdisciplinarity, the integration with the AI ecosystem, and the potential outputs.*

---

Thank you, we share the excitement and are eager to deliver on the promise of the HAVA-lab proposal.

---

*However, before proceeding to an award, the committee outlined some key comments that should be addressed. This should be done a response/annex to the proposal which will be subsequently reviewed by the committee. Note you do not need to update the proposal but just provide responses to the questions.*

*The committee's question:*

*The structure is formulated as an ICAI lab. While interesting, can you clarify how to more deeply embed in the DSC and connect to Institute for Advanced Studies. Can the structure be formulated with such a deeper embedding in mind including integration with the capabilities and features provided by these cross faculty institutes?*

---

The cross-faculty nature of the DSC and the Institute for Advanced Studies is perfectly compatible with the HAVA-lab, and we will need their community, expertise and support in realizing our ambitions. Indeed, the organizational structure of the HAVA-lab is (partly) inspired by the ICAI-lab model. From the ICAI-labs we adopt i) concentration of talent on a shared research agenda in a single lab space, ii) a tenure-track level lab manager who takes care of daily operations and supervision, and iii) an active strategy for knowledge transfer to our stakeholders. Different from ICAI-labs, which are predominantly technology-focused, HAVA-lab brings the full academic width and responsible societal use that AI at the University of Amsterdam, and the Amsterdam AI ecosystem in general, aims to offer. In this sense, the HAVA-lab is also reminiscent of the ELSA-lab formula, as used in the AI, Media & Democracy lab. It is our ambition to make the HAVA-lab a combination that absorbs the best of both the ICAI-lab and ELSA-lab ‘worlds’ and embed it in the DSC.

---

*Currently, the location is noted as a “collaborative space” for 4 days a week and the embedding appears to be within the VIS lab. Can this be at the DSC for 3-to-4 days a week? Can you comment on such a shift for the project? How would you organize this?*

---

You are right, we did not explicitly specify the HAVA-lab location in the proposal, mostly because we were not sure whether the DSC could offer such a lab space. The main reason for locating PhD candidates at the HAVA-lab is that they have a collaborative space where they can jointly spend 3-to-4 days a week. By collaborating we expect that they will discuss their topics from an interdisciplinary perspective from the start. The remaining time should be spent at the location(s) of the two co-supervisors. In terms of organization, the lab manager spends at least 2 days a week at the lab location, and the PI and Co-PI will visit at least once a week. We will stimulate the DSC (or possibly the IAS) to provide a location for such a collaborative space and associated events.

---

*The committee wondered about the inclusion of other senior AI-Ethics oriented PIs, examples could be Prof. Sennay Ghebreab, Dr. Marjolein Lanzing. Can you comment on such an addition?*

---

It is expected that the HAVA-lab research will act as a multiplier for even more intense collaboration on human-aligned video-AI across UvA-disciplines. Both Prof Sennay Ghebreab and Dr Marjolein Lanzing are excellent suggestions. In our team also the names of Dr Paula Helm and Prof Corrette Ploem surfaced as possible candidates to strengthen the AI-ethics dimension. We have discussed internally how to best leverage the broad AI-expertise within UvA, without complicating PhD supervision with more than 2 (daily) supervisors with mixed expertise. Our proposal is to establish an advisory board to which we report on a regular basis, for example by a quarterly reporting cycle and a bi-annual board meeting, and who can provide general feedback on the overall direction of the lab as well as the individual PhD projects. Beyond the ethical dimension, this advisory board should also govern the social, legal, and technical dimensions of human-aligned video-AI as we consider them all equally important for the labs' success. We are of course open to discuss with the DSC leadership the establishment of such an advisory board, its members, and its reporting cycles.

---

*The inclusion of all 7 faculties is excellent. Why are all these faculties important to the project?*

---

We strongly believe that video-AI can have a broad impact across academic disciplines, as long as it aligns with human and societal values and fits the research question(s) within the academic horizon and ambitions. The fact that we have already found Co-PI's from all seven faculties of the University of Amsterdam is a testament to our vision. All seven faculties are crucial, as they provide expertise on the different axes of human and societal values alignment, inspire us with real-world problems where human-aligned video-AI is urgent, and they are our target audience for democratization of human-aligned video-AI. As a concrete example, almost everything we know about crime is based on self-reported measures or on official registrations, known for biases like memory failure, social desirability answers, and dark number problems. With video based observations, we can start to observe actual incidences of crime, and as such measure crime in less biased ways. The potential of video based studies in criminology is illustrated by a shoplifting study showing that it was not young men of color, who were most likely to steal, as official registrations suggested and criminologists believed for a long time, but rather white middle aged women. They just never got caught or prosecuted. However, to develop video-AI tools to detect shoplifting, it is not only necessary to align with criminologists who can distinguish and define common shoplifting behavior, but it also requires involvement of legal and ethical scholars to study what determines societal acceptance of such video surveillance. Beyond the opportunity, establishing a lab with such a mixed expertise is also a risk. To assure feasibility of the PhD projects, we have each PhD student supervised by two daily supervisors, one supervisor with faculty-specific 'human-aligned' knowledge and one supervisor with 'video-AI' knowledge. Below, we detail the supervision-organization of each PhD project.

Project	Human-Aligned supervisor	Video-AI supervisor
PhD 1: Cognitive-alignment	Iris Groen	FNWI
PhD 2: Ethical-alignment	Tobias Blanke	FGw
PhD 3: Legal-alignment	Heleen Janssen	FdR
PhD 4: Criminal behavior	Marie Lindegaard	FGM
PhD 5: Skill assessment	Marlies Schijven	AUMC
PhD 6: Diagnostic training	Erwin Berkhout	ACTA
PhD 7: Responsible marketing	Stevan Rudinac	FEE
		Pascal Mettes
		FNWI

All projects and faculties play a crucial role in Objective 3 of the project, where human-aligned video-AI will be democratized throughout UvA via the DSC and its monthly meetings, its workshops and conferences. As such we hope to provide even more energy to the DSC.

---

*What will we see in 10 years from the project? How will the UvA / DSC stand out?*

---

Research on video-AI technologies and on AI alignment is currently mostly developed in isolation. The DSC Lighthouse offers the opportunity to intensively collaborate on this urgent intersection. We expect that the UvA and its DSC will be at the fore-front both in terms of human-aligned video AI technology as well by its real-world implementation, for example through advising policy makers and government bodies on deploying video-AI and smart cameras in public spaces. It is expected the lab will result in new interdisciplinary collaborations, a promising one is social science research, where human-aligned video-AI provides the means to ask challenging questions motivated by social theory and to verify them at an unprecedented experimental scale. If we succeed, the HAVA-lab will not only have national impact but may also spur interest for deployment outside of the Netherlands, in the EU, and beyond. Our ambition is that the DSC becomes the go-to-place to ask for advice on human-aligned video-AI. Last but not least, there is also potential for spin-off as the expertise we develop is as of yet unique and highly needed for responsible use of video-AI in society.

---

*How will this project support the DSC?*

---

The HAVA-lab supports the DSC by bringing together our university's talent to address the challenge of human-aligned video-AI by the advancement of data science methods and by combining it with the knowledge present around video-AI and its opportunities, pitfalls and concerns. As such, it fits the university-theme of responsible digital transformations. We further enrich the portfolio of data-driven research provided by the DSC to include video-AI and associated techniques from computer vision and machine learning. Video is generally missing from data science practices, as analysis is typically focussed on text and/or images. What is more, we offer a broad embedding in all faculties of the University of Amsterdam and the Amsterdam AI ecosystem. Last but not least, the visual nature of our research also makes for excellent DSC showcases via visuals, videos and demonstrators.

---

*Will the dashboard also provide video demonstrators and outward accessible examples?*

---

Yes. The central role of videos in the proposed research allows us to enrich the dashboard with video demonstrators. Example demonstrators include qualitative outputs of research and visual summaries of proposed human alignment along the cognitive, ethical, and legal axes. We also plan to present published research into short explainer videos that will be made publicly available on the dashboard.

---

*We will have evaluation moments for the project. What would you see as useful evaluation?*

---

Our proposal is to establish an advisory board to which we report on a regular basis, for example by a quarterly reporting cycle and a bi-annual board meeting, and who can provide general feedback on the overall direction of the lab, progress of individual PhD projects, and provide valuable connections to the broader human-aligned AI community, its developments and use cases at UvA and beyond. Besides evaluation, and being integral to the DSC, we expect constant interaction with the DSC community, which should steer us in the right direction.

---

We thank the review committee for their encouragements and their constructive feedback.