PROGRESS REPORT

RESTART 2016-2020 Programme for Research, Technological Development and Innovation

RESEARCH AND INNOVATION FOUNDATION

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης

Κυπριακή Δημοκρατία

Διαρθρωτικά Ταμεία της Ευρωπαϊκής Ένωσης στην Κύπρο

A.1. GENERAL PROJECT INFORMATION	
Project Protocol Number:	EXCELLENCE/0918/0086
ESIF Number:	**861**
Project Title:	DEtecting Stereotypes in human ComputAtioN Tasks
Project Title in Greek:	Ανιχνεύοντας Στερεότυπα στις Εργασίες Ανθρώπινης Υπολογιστικής
Host Organisation:	CYENS CoE (formerly RISE)

A.2. DESCRIPTION OF THE WORK CARRIED OUT BY ALL BENEFICIARIES DURING THIS REPORTING PERIOD.

High-level overview of the DESCANT project:

During the final months of the project, the team developed a video to communicate the key goals and achievements of the DESCANT project to the general public, which reviewers may find helpful: https://www.youtube.com/watch?v=hh4V5mUU8oM

During the second reporting period (RP2) of the project (M13 to M30), all six Work Packages (WPs) have been completed, addressing the objectives of the project, as described below. In particular, the work in RP2 focuses on three of the four objectives, as the first objective was fully addressed by the work performed in RP1.

Objective 2: **Conduct controlled experiments on popular crowdsourcing platforms to understand which workers tend to perpetuate social stereotypes, under which circumstances, and on which media artefacts.**

This objective was achieved through the work carried out in WP4 (Crowdwork Experiments) and the full set of experiments and results are presented in Deliverable 4.1 (Experimental Results).

In the Interim Report covering RP1, we described the first two experiments that were conducted, which took advantage of the "opportunity" presented to us by the COVID-19 pandemic, to study the influence of time as well as significant world events, on the behaviors of crowdworkers recruited via major online platforms. As explained in that report, this was a deviation from the planned work, but was a significant opportunity for us to study new forms of social bias in crowdwork. In RP2, we resumed work on the planned experiments described in Annex I of the DESCANT Contract. The two sets of experiments performed in RP2 are described below.

Social Bias and Crowdworker Personal Attributes. This set of experiments involved a very common task on commercial crowdsourcing platforms, in which workers are asked to describe the content of images. The target images came from the Chicago Face Database (CFD)[1], a set of 597 highly standardized, passport-style people images, which are labeled with the depicted person's demographic characteristics (e.g., age, gender, race), physical facial measurements, as well as "norming data" collected by the dataset creators (e.g., the person's level of attractiveness, trustworthiness and typicality for their respective demographic group (gender/race/age)). We used the Clickworker[2] platform to create a human-intelligence task (HIT) for the image annotation process. After a round of piloting to test various HIT set-ups, we settled on a design that first asked workers to self-report their own gender and race. They were then shown one CFD image and were asked to label the depicted person's gender (Male, Female, Other), race (Asian, Black Latino, White), as well as their level of attractiveness and trustworthiness (from 1 (not at all) to 7 (extremely)). The questions posed and the responses available were formulated as to be in parallel to the "offline" norming data provided in the CFD. Using our crowdsourced data, we were able to evaluate several sources of social bias that are predicted by social psychology, such as *in-group effects* (i.e., a worker of the same gender/race of the person depicted in the image, may yield responses that are more accurate/favorable, as compared to those provided by a worker who is not from the demographic in-group).

Social Bias and Socio-economic Considerations. This set of experiments concerned the task of labeling the content of videos, which were sourced from the Dollar Street dataset.[3] For the experiment, we selected short videos that featured the theme of "handwashing," but that were

[1] https://www.chicagofaces.org/

[2] https://www.clickworker.com/

[3] https://www.gapminder.org/dollar-street

recorded in different households around the world, and for which the monthly household income (reported in USD) was provided. Thus, the primary goal of the study was to document if and how crowdworkers' biases or stereotypical beliefs are reflected when they are asked to describe/annotate content that is diverse with respect to socio-economic status (of the household setting and/or depicted persons). Through the analyses, we observed both regional and income-level biases as reflected in the sentiments of the language used (i.e., descriptive tags) to label the videos, as well as in post-task questions (e.g., "Is this video appropriate for a public service announcement on handwashing hygiene?")

Objective 3: **Illustrate and remedy within-system the impact of social biases in crowd data in a demo image analysis system.**

Objective 3 was addressed by the work carried out in WP5 (Demonstration System), which made extensive use of the datasets collected from the *Social Bias and Crowdworker Personal Attributes* experiment. In particular, D5.2 presents a detailed analysis in which the dataset was used to train a number of classification models, which predict a person's gender, race, attractiveness and trustworthiness, given an input CFD image. In these experiments, we explored the impact of the make-up of the dataset (based on workers' personal attributes) on the behaviors of the resulting models. Models were trained on subsets of the data and their performances were compared. For instance, in predicting the gender of an individual depicted in a CFD image, we compared the prediction accuracy of the model trained on the data collected from workers who self-reported as being male, to that trained on data collected from females.

We did not attempt to remedy within-system the impact of data bias, as a number of ethical issues arose. For instance, an obvious means to do this might be to identify the demographic profile of workers who provide the most accurate data, to recruit them into completing HITs. However, this can easily be viewed as a form of employment discrimination and, given that most workers are making an honest attempt to complete the task, and that many of them engage in online crowdwork to make a living, we did not want to punish them. Thus, we opted for developing a demo image analysis system, which aims to raise awareness of social bias in crowdsourced training data. Our demo system, called RECANT, is presented in D5.3, and is available at: https://recant.cyens.org.cy/

Objective 4: Raise stakeholders' awareness on human computation and social biases in crowdsourcing

The final objective was addressed by the work implemented in WP6 (Consultation with Stakeholders). The Practitioner Guidelines we developed are motivated and outlined in D6.1. The notes and materials generated via the stakeholder engagement activities are documented in D6.2.

After completing the deliverables for WP5, and discussing the composite results from the publications produced during the DESCANT project, the team brainstormed to distill the most useful – and practical – "takeaways" from our research, in terms of mitigating social biases in crowdsourced training datasets. After discussion, we decided that it would be useful to develop two sets of guidelines, one for machine learning practitioners who are developing datasets (i.e., directly using crowdsourcing methodologies), and another who tend to reuse datasets produced by others, via crowdsourcing.

Deviations:

As will be explained in WP1, due to delays in the project's implementation, an extension of six months was requested and granted, such that DESCANT was implemented over the duration of 30 months.

A.3. EXPLANATION OF THE WORK CARRIED OUT PER WORK PACKAGE (WP).					
Work Package Number:	**1**	**Start Month:**	1	**End Month:**	30
Work Package Title	**Project Management**				
Work Package Leader	**J. Otterbacher (CYENS)**				
Partner Role	**HO**	**PA1**	**FRO**		
Person Months	**3**	**1.0**	**0.1**		

Work Package Objectives as described in Annex I of the Contract.

The purpose of this WP is to ensure the efficient, open and transparent project coordination and management of the DESCANT project by the Coordinator according to the project work plan. The Key Objectives of the WP include the:

- Provision of an effective overall management function and structure to ensure that the Coordinator and international partner fulfil their obligations within the project, with respect to the Grant Agreement;
- Production of reports, organization of meetings, and liaison with our international partner;
- Monitoring/management of project finances;
- Meeting legal and contractual obligations – updating Grant Agreement where necessary;
- Ensuring appropriate ethics are taken into consideration with regard to the user data carried out within the project;
- Facilitation of knowledge sharing among all parties involved, effective liaison with WP2: Dissemination Activities, to ensure effective exploitation, dissemination and sustainability of DESCANT approach.

Work Description and Key Results

Like in RP1, many challenges were encountered in RP2 due to the COVID-19 pandemic. There were periods of time in which it was necessary to work from home exclusively, but the team settled into a "hybrid" style of work during the final year of the project. The team's communication and coordination practices were well adjusted to the situation by RP2, as described below. In addition, some staffing issues were encountered, with one team member going on maternity leave, and another one departing from the CYENS CoE. Overall, the project met its objectives, despite some delays, as will be described.

T1.1: Project Management and Coordination (Leader: J. Otterbacher, Contributors: All)

Time management and staffing. In RP2, it was necessary to make some adjustments with respect to time management and staffing. In particular, the primary researcher employed full-time on the DESCANT project, Dr. Christoforou, took a maternity leave (8/21 – 1/22), while the researcher who was meant to lead on WP5, Dr. Partaourides, left his position at CYENS. These changes are documented in Section (Deviations).

Reporting and communication with RIF: The coordinator, with support from the team members and the CYENS Research Office, led the preparation of the final progress report, covering the project's activities from M13 to M30. The coordinator also managed the process of the grant amendment, requesting a project extension of six months due to the delays encountered as a

result of the ongoing pandemic, as well as the staffing issues. The request was approved on 28 June 2021, resulting in a revised duration of 30 months and a new end date of 3 July 2022.

T1.2: Financial management (Leader: J. Otterbacher Contributors: S. Kleanthous)

Hiring of staff.

In RP2, two researchers were hired to support WP5. In particular, A. Kafkalias was hired for six months on a 50% part-time basis, to develop and evaluate the machine learning models (T5.2, T5.3, T5.4). A second researcher, P. Pericleous, was hired for four months on a 50% part-time basis, to support the development of the demonstration system (T5.3). In addition S.Herodotou, an intern at CYENS during summer 2021 carried out work for training ML models for the needs of the project. However, the cost of Mr. Herodotou's internship was covered by CYENS own funds.

Acquisition of equipment and services. In RP2, additional services on crowdworking platforms were acquired to support the ongoing work in WPs 4 and 5.

Travel arrangements. DESCANT researchers continued to participate actively in the dissemination of results via scientific conferences. However, like in RP1, most of these conferences took place in an online (i.e., virtual) format. Two physical conference trips took place: i) AAAI ICWSM took place in M30, with researcher P. Barlas traveling to Atlanta, USA to present a full paper by the team ii) the IFIP AIAI also took place in M30, with researcher A. Kafkalias traveling to Crete, GREECE to present a full paper related to the project.

Preparation for financial reporting. The coordinator assisted the research accounting staff of CYENS and OUC in the auditing of the midterm financial report, and the related requests for additional information. She also assisted in the preparation of the final financial report of CYENS and OUC and the preparation of timesheets.

T1.3: Regular Meetings / Project Communication (Leader: J. Otterbacher, Contributors: All)

Team communication. In RP2, the DESCANT team continued to meet monthly in a regularly scheduled teleconferencing call. Minutes were taken and kept in the team's Google Drive for easy reference by all members (see D1.5 for updated meeting minutes archive). Slack continued to be extensively used for daily, simultaneous communication.

T1.4 Risk Management (Leader: J. Otterbacher, Contributors: All)

The Risk Management Plan (Section 5 of D1.2) provided a reference for dealing with issues encountered. Two of the risks outlined in the plan materialized during RP2 - #1 (Staffing issues) and #9 (Delay in implementing the demonstration system). As mentioned, to mitigate the impact of #1, additional effort was contributed by other team members; where needed, part-time help was also brought on board. To address #9, as described in the plan, we involved our students in providing initial feedback on the system design, in order to jump-start the process without having to involve industry stakeholders at a premature stage. There was intense work on the demo system in the last four months of the project, with additional meetings scheduled as needed.

T1.5 Data Management and Knowledge Production (Leader: J. Otterbacher, Contributors: All)

In RP2, the project's Data and IPR Management Plan (D1.5) continued to guide the manner in which we shared the results of our research as well as the datasets produced.

Deliverables

The following deliverables have been uploaded to the IRIS portal in RP2:

D1.5 Meeting Minutes Archive (updated with minutes from project meetings in RP2). (M30)

D1.8 RPF final report. (M30)

Work Package Number:	**2**	**Start Month:**	1	**End Month:**	30
Work Package Title	**Dissemination Activities**				
Work Package Leader	**E. Christoforou (CYENS)**				
Partner Role	**HO**	**PA1**	**FRO**		
Person Months	3	1	0.2		

Work Package Objectives as described in Annex I of the Contract.

WP2 focuses on the dissemination and exploitation of the results of the DESCANT project. Its specific objectives are to:

- Share research findings via the tried and tested modes of academic and scientific dissemination. Specifically, the team aims to target top-tier international conferences related to crowdsourcing and human computation (e.g., AAAI HCOMP) and human factors in computing (e.g., ACM CHI), as well as international peer-reviewed journals (e.g., JASIST, ACM TOIT).
- Promote access to the demonstration system: the integrated components will be packaged with associated promotional materials.
- Promote the Practitioner Guidelines: the guide will be distributed widely, drawing on the relationships of CYENS with the innovation / start-up community in the Eastern Mediterranean and Middle East.

Work Description and Expected Key Results

With respect to dissemination activities in RP2, the dissemination of the DESCANT activities and results to the scientific community continued to go smoothly. However, dissemination to the public faced a number of challenges. During the last six months of the project, the CYENS CoE faced some staff turnover, losing its communications and marketing manager. As will be explained, the team made some adjustments to its communication strategy to mitigate the negative impact.

In Spring 2022, the Transparency in Algorithms Group (TAG MRG) at CYENS, was renamed to Fairness and Ethics in AI-Human Interaction (fAIre MRG). This change was made to reflect that the group is expanding, with the promotion of Dr. S. Kleanthous to the position of co-leader.

T2.1 Web/Social Media Presences (Leader: fAIre MRG, Contributors: All)

In Spring 2022, the coordinating team (fAIre MRG at CYENS) launched its own website, in order to have full control over our web presence, and to be able to update content more frequently without having to go through CYENS marketing and communications team which, as mentioned, has experienced staffing challenges. The new website is: https://faire.cyens.org.cy/ .
Furthermore, a DESCANT project page is featured on the fAIre website: https://faire.cyens.org.cy/projects/descant/

The appropriate acknowledgements (and images) have been put into place, as well as links to the DESCANT publications repository at Zenodo, and our RECANT demonstration tool.

T2.2 Press releases and media contact (Leader: fAIre MRG, Contributors: All)

CYENS, as the host organization of the project, and OUC as PA1, each created a press release for the project conclusion. The social media channels of both institutions were exploited to further distribute news related to the project, engaging the media indirectly. As for direct media engagement, the Coordinator appeared on two local news channels (Sigma TV «Μεσημέρι και Κάτι» 17/11/2021, Omega TV "Ενημέρωση Τώρα" 1/12/2021), as well as two radio shows (Active Radio 16/12/2021, CyBC "Saskia Unreserved" 20/1/2022), and one internet news outlet based in the UK (The News Movement 6/2022.) These opportunities were used to raise the public's awareness of algorithmic bias, and the persistent issue of biased data.

T2.3 Scientific publications (Leader: J. Otterbacher, Contributors: All)

During RP2, 10 scientific publications were produced in double-blind peer-reviewed venues:

- Barlas, P., Krahn, M., Kleanthous, S., Kyriakou, K., & Otterbacher, J. (2022, May). Shifting Our Awareness, Taking Back Tags: Temporal Changes in Computer Vision Services' Social Behaviors. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 22-31).
- *Fan, S., Barlas, P., Christoforou, E., Otterbacher, J., Sadiq, S., and Demartini, G. (2022). Socio-Economic Diversity in Human Annotations. In 14th ACM Web Science Conference 2022 (WebSci '22). Association for Computing Machinery, New York, NY, USA, (pp. 98–109). https://doi.org/10.1145/3501247.3531588*
- Kafkalias, A., Herodotou, S., Theodosiou, Z., & Lanitis, A. (2022). Bias in Face Image Classification Machine Learning Models: The Impact of Annotator's Gender and Race. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 89-100). Springer, Cham.
- Kasinidou, M., Kleanthous, S., & Otterbacher, J. (2021, July). 'Expected Most of the Results, but Some Others... Surprised Me': Personality Inference in Image Tagging Services. In *International Symposium on End User Development* (pp. 187-195). Springer, Cham.
- Kleanthous, S., Kasinidou, M., Barlas, P., & Otterbacher, J. (2022). Perception of fairness in algorithmic decisions: Future developers' perspective. *Patterns*, *3*(1), 100380.
- Kyriakou, K., Barlas, P., Kleanthous, S., Christoforou, E., & Otterbacher, J. (2021). Crowdsourcing Human Oversight on Image Tagging Algorithms: An initial study of image diversity. In Ninth AAAI Conference on Human Computation and Crowdsourcing Work-in-Progress (HCOMP 2021).
- Orphanou, K., Otterbacher, J., Kleanthous, S., Batsuren, K., Giunchiglia, F., Bogina, V., Shulner Tal, A., Hartman, A., and Kuflik, T.. 2022. Mitigating Bias in Algorithmic Systems - A Fish-Eye View. *ACM Computing Surveys. Just Accepted (March 2022).*
- Orphanou, K., Christoforou, E., Otterbacher, J., Paramita, M. L., & Hopfgartner, F. (2021, November). Preserving the memory of the first wave of COVID-19 pandemic: Crowdsourcing a collection of image search queries. In *Proceedings of the Third symposium on Biases in Human Computation and Crowdsourcing*. CEUR Workshop Proceedings.
- Paramita, M. L., Orphanou, K., Christoforou, E., Otterbacher, J., & Hopfgartner, F. (2021). Do you see what I see? Images of the COVID-19 pandemic through the lens of Google. *Information Processing & Management*, *58*(5), 102654.
- Perikleous, P., Kafkalias, A., Theodosiou, Z., Barlas, P., Christoforou, E., Otterbacher, J. Demartini, G. and Lanitis, A. How Does the Crowd Impact the Model? A tool for raising awareness of social bias in crowdsourced training data. To appear in the Proceedings of

31st ACM International Conference on Information and Knowledge Management (CIKM'22), November 2022.

One additional submission is still under review:

- Barlas, P., Krahn, M., Kleanthous, S., Kyriakou, K., and Otterbacher, J. Social B(eye)as Over Time: Dataset for temporal auditing of image tagging algorithms. Submitted to the 36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks.

T2.4 Guidelines for practitioners (Leader: fAIre MRG, Contributors: All)

In the context of WP6, the team developed a set of easy-to-read guidelines, which aim to communicate the practical implications and tips that come out of our research in the DESCANT project. As will be described in WP6, one guideline was developed for practitioners who actually perform crowdsourcing in the process of developing training datasets, while another focused on tips for those who do not directly use crowdsourcing, but who (re)use the crowdsourced datasets created by others. In addition to the narrative provided in D6.1, we have also developed an infographic version for each of the guidelines.

T2.5 DESCANT Seminar(s) (Leader: fAIre MRG, Contributors: All)

During RP2, the following *invited talks / seminars / keynotes* were given in Cyprus and abroad, during which the DESCANT project was highlighted:

- P. Barlas, E. Christoforou, K. Kyriakou, S. Kleanthous & J. Otterbacher. *A critical look at Image Tagging Algorithms*. Austria Critical Data Seminar, University of Art and Design Linz, March 2021
- J. Otterbacher. *It's about time…and perspective: A critical look at the use of crowdsourcing in building image datasets.* Delft University of Technology, Web Information Systems' research group, lecture series on "Responsible Use of Data", May 2021.
- J. Otterbacher. *What is Beautiful Continues to be Good: Images of Women and Algorithmic Inferences on Identity*. POWER Workshop (in Greek), May 2021.
- S. Kleanthous & J. Otterbacher. *A critical look at the use of crowdsourcing in building image datasets*. IBM Benelux Data & AI Ethics seminar series, June 2021.
- J. Otterbacher. *A critical look at the use of crowdsourcing in building image datasets.* U.S. Naval Research Lab AI Center's Virtual Seminar Series, June 2021.
- J. Otterbacher. *Open, Fair and (Un)biased Data: Lessons from computer vision and beyond. CLIPE – MSCA-ITN Training School, September, 2021.*
- J. Otterbacher. *Computer Vision in the Social World: On fairness and fairness perception in image tagging algorithms.* International Conference on Computer Analysis of Images and Patterns (CAIP 2021), September 2021.
- S. Kleanthous. *AI Biases and Inclusion*. Invited Talk at IEEE Digital Reality Series - Future Directions, October 2021.
- J. Otterbacher. *Big Data and AI: Challenges for Education*. Invited Keynote talk at the Cyprus Pedagogical Institute's Annual Conference on ICT for Education, November 2021.
- S. Kleanthous. *AI Biases and Inclusion*. Invited talk at the University of Trento, KnowDive Research Group seminar series, November 2021.
- J. Otterbacher. It's about Time....and Perspective - *A critical look at proprietary computer vision algorithms and the data practices behind them*. Invited session at the 1st Greek ACM-W Chapter Winter School, Fairness in AI, February 2022.

- J. Otterbacher. *It's about Time....and Perspective - A critical look at proprietary computer vision algorithms and the data practices behind them.* Nokia – Bell Labs Cambridge, Seminar on Responsible AI, March 2022.
- S. Kleanthous & J. Otterbacher. *Artificial Intelligence in Everyday Life: A discussion on its promises and limitations.* Cyprus Third Age (C3A) seminar, June 2022.
- J. Otterbacher. *Algorithmic Analysis of People Images - Reflections on fairness, accuracy, and the use of computer vision for security and privacy applications.* Keynote talk at APPS Workshop, held in conjunction with the ACM UMAP conference, in Barcelona, July 2022.

In addition to the above, two keynote talks at international conferences have been scheduled for J. Otterbacher and will discuss some of the DESCANT work (14th International Conference on Computational Collective Intelligence (ICCCI), September 2022; International Conference of Photography and Theory, November 2022).

Deliverables

The publications resulting from DESCANT can be found in the project's Zenodo repository: https://zenodo.org/communities/descant_cy

The following deliverables have been uploaded to the IRIS portal in RP2:

D2.3 Project publication list and archive (Month 30).

D2.4 Publicly accessible archive of all materials and demonstration system (Month 30).

Work Package 3, *Conceptual Framework*, concluded in RP1.

Work Package Number:	4	Start Month:	5	End Month:	24
Work Package Title	Crowdwork Experiments				
Work Package Leader	G. Demartini (UQ)				
Partner Role	HO	PA1	FRO		
Person Months	6.0	1.0	1.5		

Work Package Objectives as described in Annex I of the Contract.

WP4 focuses on achieving an understanding of how and when socially biased responses to human intelligent tasks are observed. This WP carries out controlled experiments on two different commercial crowdsourcing platforms based on the results of WP3.

Its key objective is to:

- Uncover the correlations between the characteristics of the HIT, the worker characteristics, and the characteristics of the respective multimedia artefact, and the expression of social stereotypes in the data produced.

Work Description and Expected Key Results

T4.1 Experiments on MTurk and FigureEight (Leader: G. Demartini, Contributors: All)

In total, a series of four experiments was carried out; experiments are described in detail in D4.1. Two were launched in RP1 and two were conducted in RP2. The table below presents the key experimental parameters under consideration across the four studies.

	Social Bias and Personal Attributes	**Time and Event-Related Effects**	**Social Bias and Socio-Economic Status (SES)**	**Cultural and Geographical Bias**
Bias explored	Bias and stereotypical beliefs correlated to personal characteristics	Temporal effects of significant global and local events (e.g., COVID-19, social unrest in the U.S.)	Social stereotypes as a result of differences in the SES of workers / depicted persons	Cultural and geographical bias in creating a dataset relevant to COVID-19 images
Media artefact	People-related images	People-related images	Videos depicting people washing their hands	Text, presentation of situational/conditional story/scenario
Platform	Clickworkers	Appen	Amazon Mechanical Turk	Clickworkers
HIT key characteristic(s)	Closed-form responses	Open-form responses (tags)	"Survey" on handwashing activity depicted in videos, from different geographical regions and SES	Collection of image search queries
Worker key characteristic(s)	U.S. residents, self-reported race and gender	U.S. residents, self-reported race and gender	U.S residents	Residents of Italy, Spain, Great Britain and Germany

Bias identification method	Quantify the effect that data from workers with different characteristics have on the performance of machine learning classification models	Replication of a study performed in 2018; comparison with historic data	Statistical analysis of tags, categorization of tags and consideration of diversity	Result analysis with thematic categorization of responses

T4.2 Data analysis (Leader: G. Demartini, Contributors: All)

In RP2, all the data collected across the four experiments were analyzed (guided by the framework developed in RP1 and that is presented in D3.3 – *Description of statistical tests and procedures*) and the results were published.

Deliverables

D4.1 Experimental Results. (M14).

Work Package Number:	5	Start Month:	15	End Month:	30
Work Package Title	Demonstration System				
Work Package Leader	A. Lanitis (CYENS – BIO-SCENT MRG)				
Partner Role	HO	PA1	FRO		
Person Months	8.0	1.0	0.1		

Work Package Objectives as described in Annex I of the Contract.

The goal of WP5 is to *tangibly demonstrate* how the underlying social biases in training data can affect the performance of an algorithmic system that is trained on that data. For this purpose, we focus on the task of machine learning for automatic content tagging on (i.e., the generation of descriptive metadata) of images depicting people. The images will have been labelled by crowdworkers in the context of the experiments performed in WP4, such that the researchers will be fully aware of the circumstances under which the data was produced (i.e., the entire HIT setup on MTurk and/or FigureEight).

The key objectives of WP5 are to:
1) Demonstrate how social biases in the underlying human intelligence data affects system performance in a prototypical artificial intelligence-based application
2) Create a user-friendly interface to the system, such that it can be used for training purposes.
3) Evaluate the effect of transfer learning to deal with social biases.

Work Description and Expected Key Results

T5.1 Development of data sets (Leader: fAIre MRG, Contributors: All)

This task involved the data collection, cleaning and archiving of the datasets resulting from the experiments described in WP4. Deliverable 5.1 (Training Datasets) describes these processes in detail.

T5.2 Development of image analysis algorithms (Leader: BIO-SCENT MRG, Contributors: All)

The dataset on *Social Bias and Personal Attributes* was created with the development of our demonstration system in mind. As explained in D4.1 and D5.2, the dataset features several biometrics tasks – given a passport-style image of a person. Two main sets of experiments were performed, using ML models trained using Python and Tensorflow, and using ML models trained using a publicly available software platform.

In the first set of experiments the tasks of attractiveness, trustworthiness, age, gender and race estimation were considered. For the second set the tasks of estimating the depicted person's gender, race, perceived level of attractiveness, and perceived level of trustworthiness are considered. Although the tasks of gender and race estimation represent rather "standard" biometrics tasks, the tasks of attractiveness, and perceived level of trustworthiness estimation are likely to be seen as more esoteric. However, these sorts of algorithms are now being used in many applications, such as those used for screening job applicants (e.g., products such as HireVue or Affectiva). Our dataset was built using the Chicago Face Database, which contains 597 standardized people images, along with norming data, across many attributes such as demographics (e.g., gender, age, race) and "person perception" aspects (e.g., attractiveness,

trustworthiness, typicality). Thus, we have data on many attributes from the Database creators, as well as responses from crowdworkers on the above mentioned four attributes; furthermore, workers self-reported their own gender and race.

Using this data, machine learning experiments were conducted, resulting in classification models for the biometric tasks considered. D5.2 explores in detail the performance of each model, as well as the performance of models that are trained on subsets of the data, with respect to worker attributes (e.g., the model trained on data from all workers identifying as men). Thus, the experiments explore the relationship between the diversity of the crowd data, and model performance.

T5.3 Development of demonstration system (Leader: BIO-SCENT MRG, Contributors: All)

A Web-based demonstration tool, RECANT, was created in this task, based on the classification models developed in T5.2. The tool is designed to raise awareness of the crucial dependency of model performance on the properties of the training dataset. Specifically, the demonstration tool allows users to explore the classification decisions of the various models, for an image of their choice (i.e., from a subset of images taken from the Chicago Face Database). The details of the system development can be found in D5.3.

T5.4 Transfer learning development (Leader: BIO-SCENT MRG, Contributors: All)

This last task was implemented in a slightly different way than originally envisioned. In the spirit of "AI Democratization," the model training was done using the lobe.ai tool, which makes use of open-source machine learning architectures. The tool has the advantage of ensuring that all models under comparison were trained using the exact same training and model optimization procedures. One disadvantage of lobe.ai is that it does not provide explicit details of the model architectures or the training algorithms that it uses. However, this procedure represents how modern AI developers work, using online tools that are efficient, effective and cost efficient. This process is described in detail in Deliverable D5.2.

Deliverables

D5.1 Datasets. (M16).

D5.2 Analysis of algorithm performance in various settings. (M20).

D5.3 Fully implemented demonstration system. (M30)

Work Package Number:	6	Start Month:	18	End Month:	30
Work Package Title	Practitioner Guidelines				
Work Package Leader	S. Kleanthous (OUC)				
Partner Role	HO	PA1	FRO		
Person Months	2.0	8.0	0.1		

Work Package Objectives as described in Annex I of the Contract.

WP6 focuses on generating impact by creating tangible outcomes that can foster innovation in the area of human-in-the-loop system design. Building on top of the results achieved in WP3-5, in WP6 we will conduct activities that raise stakeholders' awareness on the presence of social bias in manually labeled datasets. The goal is to understand the practical challenges faced by stakeholders in order to best "translate" the research results produced in WP3-5 into practical insights for practitioners, concerning how to best design crowdsourcing tasks to minimize biases.

Its key objectives are to:

1) Conduct consultations with practitioners who are using crowdsourcing in their work, in order to understand their needs and the challenges they are facing.

2) Produce practical and easy-to-use guidelines, that will aid practitioners in dealing with the issue of social biases in the data they collect via paid crowdsourcing.

3) Solicit feedback from stakeholders on our guidelines as well as our demonstration system.

Work Description and Expected Key Results

T6.1 Draft Practical Guidelines (Leader: CyCAT, Contributors: All)

In this task, the team reviewed the survey of the literature produced in RP1 (D3.1) as well as the publications produced by the DESCANT team. The goal was to distill the set of key "take-aways" that would be most useful to practitioners, in terms of mitigating social and cultural bias in crowdsourcing and/or crowdsourced datasets that they are using. Through this process, a draft guideline document was produced, which first motivated the need for the guidelines and then outlined the key tips.

T6.2 Consultations with Stakeholders (Leader: CyCAT, Contributors: All)

Two roundtable discussions were organized with local stakeholders. In particular, we relied on the CYENS network to identify and reach out to machine learning practitioners who are either directly or indirectly involved in crowdsourcing. D6.2 describes the outline of our events, as well as information about the stakeholders. It also presents the materials we developed for these events.

T6.3 Revised Practical Guidelines (Leader: CyCAT, Contributors: All)

Based on the feedback received during the roundtable discussions, we slightly refined the practitioner guidelines. The insights gained from the events are discussed in D6.2.

Deliverables

D6.1 Practitioner Guidelines. (M30).

D6.2 Notes and materials generated from the consultations with stakeholders. (M30).

A.4. TABLE OF WORK PACKAGES

		Contract		Actual Implementation	
Work Package	**Work Package Title**	**Start Month**	**End Month**	**Start Month**	**End Month**
WP1	**Project Management**	1	24	1	30
WP2	**Dissemination Activities**	1	24	1	30
WP3	**Conceptual Framework**	1	4	1	5
WP4	**Crowdwork Experiments**	5	14	6	24
WP5	**Demonstration System**	15	20	15	30
WP6	**Practitioner Guidelines**	18	24	18	30

A.5. EXPLOITATION OF RESULTS AND ADDED VALUE

As described in Section B3.1 of the DESCANT contract, the project aims to generate added value in two key areas – i) its scientific impact, ii) its social and economic impact.

Scientific impact.
DESCANT tackled a novel research problem – the detection of socially biased responses in micro-task crowdsourcing, in which respondents (i.e, crowdworkers) are asked to annotate/judge people-related media (e.g., image tagging, demographic inferences from images). To this end, DESCANT targeted high-profile publication venues, to make impact in the research communities focusing on human computation and crowdsourcing, as well as more generally, human-computer interaction. During the 30 months of the project, the following achievements were made
:

- Publishing in high-profile international conferences, e.g., AAAI HCOMP, which is the premier international conference focused on human computation and crowdsourcing research, ACM CHI, which is the top-ranking venue for human-computer interaction research, and ACM FAccT, which is an emerging international community that is focused on issues of fairness, accountability and transparency in data and algorithms.
- One industry-related proposal was submitted to the Facebook call for faculty research grants, as a collaboration between UQ and CYENS. While the proposal was not successful, it laid the groundwork for a jointly authored paper in the ACM WebSci conference (see Fan et al., 2022, in the publications list for DESCANT).
- DESCANT provided an excellent opportunity for young researchers in Cyprus to receive high-quality training, as well as to interact with colleagues at the FRO (UQ). Solid evidence of the benefits of this training could be seen early on. For instance, Dr. Evgenia Christoforou is now successfully coordinating the KeepA(n)I project (see below for more details). Pinar Barlas has received a full scholarship for doctoral studies at Western University (Ontario, Canada) beginning in September 2022. Finally, Dr. Styliani Kleanthous received a promotion at CYENS CoE, and is now co-leader of the fAIre research group.
- OUC (CyCAT) has become a sponsor of the ongoing Biases in Human Computation and Crowdsourcing (BHCC) symposium, which brings together European researchers addressing these issues each fall for its annual event.

Societal impact.
Equally important to DESCANT is the goal of making societal impact. In particular, DESCANT aimed to promote digital entrepreneurship, particularly in the local community in Cyprus. Thus, in addition to its communication activities, the team engaged in the following measures:

- Participation in the CYENS CoE annual summer school, which is open to youths 11+ years old. In particular, the DESCANT team participated in the 2020 and 2021 camps, with interactive presentations on “everyday AI” applications (and their potential biases) featuring the team’s demos.
- Hosting of summer interns, who collaborated with our teams. For instance, in the summer of 2021 BIO-SCENT hosted an intern, Stylianos Herodotou, who developed a project on “Investigating bias in machine learning models.” Similarly, the 2021 intern with fAIRe MRG, Maximillian Krahn, collaborated in developing replication studies for previous audits conducted, and ended up co-authoring a publication with the team (see Barlas et al., 2022 in AAAI ICWSM).

- The DESCANT team has produced three motion graphics videos, for the purpose of educating the public on the issue of Data Bias, Computer Vision Bias and more generally the role of educational interventions in combating these issues. (These videos were developed in collaboration with OUC, who provided the funding for the video production.)
- In the context of WP6, two infographics were produced to explain, in an accessible way, the practical implications of DESCANT research on data practices.
- Finally, the gender dimension of the project has been emphasized all along, both in terms of the project's content and goals (i.e., fighting social biases in data and the algorithms trained on them) and the project's management and staffing. In particular, five women researchers played key roles in the project (Otterbacher – coordinator, Christoforou, Kleanthous, Barlas, Kasinidou).

Exploitation of Results

The DESCANT contract promised to use the project's results to:

- inform the research community by means of the newly generated knowledge about the presence of bias and stereotypes in the data collected from paid crowdsourcing platforms.
- showcase how tracking and surfacing stereotype information in manually labeled datasets can inform end-users about the presence of bias in digital content.

Regarding the first goal, as mentioned, a total of 15 high-impact scientific publications were produced, with more currently under review. Furthermore, as reported in WP2, a total of 14 invited presentations / seminars / keynotes were delivered by DESCANT researchers.

With respect to the second goal, our RECANT demonstration system is the main vehicle by which we have showcased the issue of data bias in crowdsourcing training data, and its potential effects on algorithmic models. We believe that the best recourse for combatting data bias is to raise the awareness of machine learning practitioners. We are already using RECANT in our own lectures and lessons with our students, and we plan to continue to develop exercises and other ways to integrate this approach into computer science / data science curricula, but also, in more informal training settings. Finally, the RECANT demo system has been selected for presentation at the prestigious ACM CIKM conference, which shall take place in November 2022 in Atlanta, Georgia USA.

Other Impacts

- New positions created (or retained)

DESCANT resulted in the creation of one full-time and one part-time position at the CYENS CoE (HO), and one full-time PhD student position at OUC (PA1). Specifically, Dr. Evgenia Christoforou was hired in M1 of the project, on a full-time basis. She continues on at CYENS, having been successful in the 2021 call of Excellence Hubs, with the follow-up project, KeepAnI. Two part-time researchers were employed on DESCANT at CYENS, to support WP5. Finally, Maria Kasinidou, a PhD student at OUC, has been employed as a research assistant and her employment continues on at OUC, also on another project.

- Collaboration with organizations abroad

DESCANT was instrumental in solidifying collaborations between CYENS, OUC and UQ. In addition to activities planned with the project, further collaboration included the submission of a

joint industry-research proposal to Facebook, as well as joint organization of workshops and conferences (e.g., the AAAI ICWSM 2023, which will take place in Cyprus and in which Dr. Gianluca Demartini will lead the organization of the tutorials, while Dr. Jahna Otterbacher serves as the General Chair).

Additional relationships were enhanced through DESCANT as well. For instance, members of fAIre team at CYENS (Otterbacher, Christoforou, Kyriakou, Kleanthous) are actively collaborating with the University of Trento (and the team of Prof. Fausto Giunchiglia) in a number of ways including co-authoring of articles, co-supervision of PhD students, as well as proposal writing. In addition, a strong collaboration with the University of Sheffield's Information School (in particular, data science faculty members Dr. Frank Hopfgartner and Dr. Monica Lestari Paramita) has developed, which has included joint publications as well as proposal writing.

- New processes and services

The research performed in DESCANT, as well as the stakeholder engagement activities, have yielded many insights as to how to support individual developers and start-up companies in their efforts to get involved with data-driven AI. The push to "democratize" AI and provide access to everyone to key components such as data, algorithms, and computing power, and at the same time, the fact that regulation (the EU AI Act) is on the horizon, create an environment in which stakeholders will need support to ensure that their products and services are trustworthy (or at least, that due diligence has been taken in this regard). The DESCANT team is working with the Research and Innovation Department at CYENS to develop consultation and auditing processes and services, which focus on understanding the sources of bias in data and algorithms, and mitigating them. In summary, ideas and insights developed in DESCANT will feed into the work to be done in the context of CYENS' new project, DiGiNN, funded under the European Digital Innovation Hubs call, which will launch in November 2022.

- Proposals submitted in H2020 and Horizon Europe

The team contributed to five European proposals during the 30 months of the project, as shown in the table below. To date, one has been funded, on which the two CYENS teams of DESCANT (BIO-SCENT, fAIre) are involved.

Title	**Funding instrument, Total value**	**Involvement of DESCANT personnel and partners**	**Outcome**
AIdvice: AI-driven Crowd-based Ecosystem with Credibility Modelling for Countering Disinformation	HORIZON-CL4-2021-HUMAN-01 6.435.971 euro	Partners. Coordinator is the University of Glasgow.	Not funded.
VIVEKA: the Power of Awareness	HORIZON-EIC-2021-PATHFINDERCHALLENGES-01 4.347.281 euro	Partners. Coordinator is the University of Trento.	Not funded.
CASCADE: CyCAT as a Facilitator of Algorithmic Oversight and Public Engagement	HORIZON-WIDERA-2021-ACCESS-03 Twinning 1.499.971 euro	Coordinators.	Not funded (but over threshold).
PROGRESS: Promoting Lifelong Learning Through Data	ERASMUS-EDU-2022-PI-FORWARD 999.999 euro	Coordinators.	Under review.

Science Skills for European Citizens.			
DiGiNN: Cyprus DIGital INNovation Hub	DIGITAL-2021-EDIH-01	CYENS is a partner on the project, which is coordinated by The Cyprus Institute.	Funded.

- Follow-up project funded by RIF Excellence Hubs

In the 2021 call for Excellence Hubs proposals, the fAIre team (with Dr. Christoforou as coordinator), was successful with the proposal “KeepA(n)I: A Methodological Approach for Identifying Social Stereotypes in Artificial Intelligence Applications.” DESCANT research laid the groundwork for the KeepA(n)I project, which will develop human oversight mechanisms for computer vision algorithms, using a crowdsourcing approach. The project is a collaboration with industry partner, AlgoLysis Ltd., and received 199.999 euro in funding.

B.1. ADDITIONAL INFORMATION (OPTIONAL)

Deviations concerning the use of resources:

Other direct costs (CYENS)
In the DESCANT contract, Other Direct Costs, two specific commercial crowdsourcing platforms are mentioned: FigureEight and Amazon Mechanical Turk. It should be noted that the FigureEight platform has undergone a number of changes, both in terms of its name (it was first renamed CrowdFlower, and is currently called Appen), as well as in terms of its ownership and business model. Specifically, Appen.com no longer accepts academic use of its platform. However, given that the coordinator (J. Otterbacher) had an existing account with FigureEight, we were allowed to execute a smaller number of crowdsourcing tasks.

Given that Appen will no longer be accessible for academic research, we have opted to use Clickworker, a platform based in Germany that is increasing in popularity both amongst academics as well as within industry. In addition, we have used the Prolific platform, which is another popular alternative that is based in the UK.

Amazon Mechanical Turk (MTurk) was used for the experiment on *Social Bias and Socio-Economic Status (SES).* However, due to problems with the manner in which MTurk (based in the U.S.) issues invoices, we were unable to pay for these services from the DESCANT funds. (Our use of MTurk was covered from another funding source.)

Personnel / Person Months (CYENS)
As noted in WP1, there were some changes in the personnel who worked on the DESCANT project. In particular:

- Dr. Haris Partaourides (CYENS – BIO-SCENT) was replaced by two part-time researchers (A. Kafkalias, P. Pericleous) to cover his contributions to WP5. This change was necessary as Dr. Partaourides left CYENS before the start of the WP.
- Dr. Evgenia Christoforou (CYENS – fAIre) was on maternity leave from 8/2021 to 12/2021. To keep the implementation moving on WP4 and WP5, additional contributions were made by J. Otterbacher and P. Barlas.

CYENS CoE (HO), has been able to contribute more PMs than originally planned, contributing a total effort of 20 PMs during the 18 months of RP2. As can be seen in the table below, significant additional efforts were invested into the dissemination and communication efforts (WP2), as well as to the crowdsourcing experiments (WP4). Slightly more effort (one PM) was put into the stakeholder engagement (WP6) as well as the development of the demonstration system (i.e., 0.5 additional PMs in WP5).

	WP1	**WP2**	**WP3**	**WP4**	**WP5**	**WP6**	**Total**
Total Planned	**3.0**	**3.0**	**4.0**	**6.0**	**8.0**	**2.0**	**26.0**
Declared – RP1	1.5	1.0	5.0	4.0	0	0	11.5
Declared – RP2	1.5	3.0	0	4.0	8.5	3.0	20.0

Personnel / Person Months (OUC)
The Open University of Cyprus (OUC - PA1), has been able to contribute more PMs than originally planned, contributing a total effort of 14 PMs during the 18 months of RP2.

Specifically, PMs were contributed by researcher Maria Kasinidou, who is a PhD student at the OUC. As can be seen in the table below, there were deviations with respect to WP2 (with two more PMs than planned) as well as WP3 (0.9 PMs more than planned). More effort was put into the efforts for dissemination activities (particularly scientific publications), in order to create more impact, and to enhance the experience of the student. In addition, more effort was invested in WP3, and in particular, in the development of the conceptual framework underpinning the project, in its early stages.

	WP1	**WP2**	**WP3**	**WP4**	**WP5**	**WP6**	**Total**
Total Planned	**1.0**	**1.0**	**0.1**	**1.0**	**1.0**	**8.0**	**12.1**
Declared – RP1	0.5	1.5	1.0	0	0	0	3.0
Declared – RP2	1.0	3.0	1.0	1.0	1.0	7.0	14.0

B.2. FOLLOW-UP OF RECOMMENDATIONS AND COMMENTS FROM PREVIOUS PROJECT EVALUATION

Not applicable.