

### Poster session

#### Friday, July 9 16:30-17:30

Poster Nb	Name	Ecole/Company	Poster title
1	Amir-Hossein BATENI	ENSAE Paris, Institut Polytechnique de Paris	Confidence regions and minimax rates in outlier-robust estimation on the probability simplex
2	-	-	-
3	Chuang YU	U2IS, ENSTA Paris, Institut Polytechnique de Paris	A one-to-many co-speech robot gesture synthesis
4	Daria MOROZOVA	HEC Paris	I, Human: How Artificial Agents Make Us Try Harder (Or Not) In Creative Tasks
5	Avetik KARAGULYAN	ENSAE Paris, Institut Polytechnique de Paris	Penalized Langevin dynamics with vanishing penalty for smooth and log-concave targets
6	Laura TINSI	CREST, ENSAE Paris, Institut Polytechnique de Paris & EDF	Risk bounds for aggregated shallow neural networks using Gaussian priors
7	Mohamed ALAMI CHEHBOUNE	LIX, Ecole Polytechnique, Institut Polytechnique de Paris	Transferable Deep Metric Learning for clustering
8	Myrto LIMNIOS	Telecom Paris, Institut Polytechnique de Paris	Concentration Inequalities for Two-Sample Rank Processes with Application to Bipartite Ranking
9	Zhegong SHANGGUAN	U2IS, ENSTA Paris, Institut Polytechnique de Paris	Driving behavior Analysis from multi-modal data
10	Maya GUILLAUMONT	Capgemini	How to assess the environmental impact of AI? This is the challenge of Sustainable AI Project
11	Olivier MATZ	Capgemini	Speech recognition under constraint

# Poster Abstracts

For more details on the posters: <https://www.summerschool.hi-paris.fr/program/>

## Poster 1 - Confidence regions and minimax rates in outlier-robust estimation on the probability simplex

**Speaker:** Amir-Hossein BATENI - [amirhossein.bateni@ensae.fr](mailto:amirhossein.bateni@ensae.fr)

**Abstract:** We consider the problem of estimating the mean of a distribution supported by the  $k$ -dimensional probability simplex in the setting where an  $\epsilon$  fraction of observations are subject to adversarial corruption. A simple particular example is the problem of estimating the distribution of a discrete random variable. Assuming that the discrete variable takes  $k$  values, the unknown parameter  $\theta$  is a  $k$ -dimensional vector belonging to the probability simplex.

We establish minimax rates when the quality of estimation is measured by the total-variation distance, the Hellinger distance, or the  $L_2$ -distance between two probability measures. We also provide confidence regions that shrink at the minimax rate. Our analysis reveals that the minimax rates associated to these three distances are all different, but they are all attained by the sample average. Furthermore, we show that the latter is adaptive to the possible sparsity of the unknown vector.

## Poster 2 -

**Speaker:**

**Abstract:**

## Poster 3 - A one-to-many co-speech robot gesture synthesis

**Speaker:** Chuang YU - [chuang.yu@ensta-paris.fr](mailto:chuang.yu@ensta-paris.fr)

**Abstract:** The human gestures occur spontaneously and usually they are aligned with speech, which leads to a natural and expressive interaction. Speech-driven gesture generation is important in order to enable a social robot to exhibit social cues and conduct a successful human-robot interaction. In this paper, the generation process involves mapping acoustic speech representation to the corresponding gestures for a humanoid robot. The paper proposes a new GAN (Generative Adversarial Network) architecture for speech to gesture generation. Instead of the fixed mapping from one speech to one gesture pattern, our end-to-end GAN structure can generate multiple mapped gestures patterns from one speech (with multiple noises) just like humans do. The generated gestures can be applied to social robots with arms. The evaluation result shows the effectiveness of our generative model for speech-driven robot gesture generation.

## Poster 4 - I, Human: How Artificial Agents Make Us Try Harder (Or Not) In Creative Tasks

**Speaker:** Daria MOROZOVA - [daria.morozova@hec.edu](mailto:daria.morozova@hec.edu)

**Abstract:** Although creativity is supposed to help humans compete against artificial agents (AAs; robots, AI), little research examines how AAs impact creative process. Dominant belief is that AAs are unsuited for creative work, considered primarily human capacity (Takayama et al., 2008). As peer effects influence effort (Zimmerman, 2003), we suppose that (H1) exposure to AAs results in decreased effort in 'uniquely human' tasks.

Consequently, AAs' creative capacity should be surprising. Expectation violation might result in contrast effect leading to increased self-maintenance efforts (Roese and Sherman, 2007). When distant comparison target turns out to be self-proximate on an important dimension, that dimension

becomes more salient (Garcia and Tor, 2007). Thus, dimension-maintenance efforts increase: (H2) exposed to a surprisingly creative AA, individuals exert more creative effort to protect human identity. We conducted four experiments (n=1396; S1 and S2 British, S3 and S4 Russian sample). In Study 1, 224 participants collaborated with AI/a human on a creative (adapted Guilford task, S1a) and a non-creative (word-search, S1b) task. AI-collaboration (vs. human) predicted shorter time taken on creative task, but no effect was observed for non-creative task.

In Study 2, 218 participants competed against AI/a human on S1 tasks. In AI-competition, AI-threatened participants who had experience with it marginally increased creative effort (p=.09). For the non-creative task, no predictors except for older age were significant.

Study 3 addressed effort in neutral exposure to counterpart performance. In creative (3a) task, 225 participants proposed ideas for a wellbeing committee, looked at appropriate/nonsense ideas suggested by human/AI (nonsense ideas really generated by an algorithm), and again proposed ideas on a different topic. Human-condition participants spent almost a half-minute longer on the second task than the AI-condition. In non-creative (3b) task, 225 participants searched for a character on an old-Russian birch bark, looked at appropriate (correct characters found)/nonsense (all 452 characters highlighted) performance of human/AI, and searched for a different character. Contrary to predictions, in AI-condition participants took on average 6 seconds fewer.

Study 4 investigated the effect of creativity uniqueness salience on creative effort in neutral AA-exposure. 264 participants primed with creativity uniqueness/not primed at all were pre-task exposed to above/below average performance of a human/AI on the same task. Creativity-primed participants exposed to high-performance AI worked on average for five minutes (overall longest), and ~1.5 mins longer than creatively-primed participants in high-performance human group. In low-performing conditions, non-primed human-exposed participants worked for ~2.5 minutes, almost the same as non-primed AI-exposed participants. When AAs' high creative performance was unexpected, perceived AA-threat increased.

AA-exposure makes people exert less creative effort due to perceived unsuitability of AAs for creative tasks. If AAs are surprisingly creative, individuals exert greater effort when creativity as a uniquely human characteristic is self-salient, and perceived AA-threat increases. Our contribution is in identification of the mechanism linking perceptions of AAs and social comparison theory. We also demonstrate the moderating role of lay AA-perceptions and their uniformity in the Western and Eastern European societies.

## Poster 5 - Penalized Langevin dynamics with vanishing penalty for smooth and log-concave targets

**Speaker:** Avetik KARAGULYAN - [avetik.karagulyan@ensae.fr](mailto:avetik.karagulyan@ensae.fr)

**Abstract:** We study the problem of sampling from a probability distribution on  $\mathbb{R}^p$  defined via a convex and smooth potential function. We first consider a continuous-time diffusion-type process, termed Penalized Langevin dynamics (PLD), the drift of which is the negative gradient of the potential plus a linear penalty that vanishes when time goes to infinity. An upper bound on the Wasserstein-2 distance between the distribution of the PLD at time  $t$  and the target is established. This upper bound highlights the influence of the speed of decay of the penalty on the accuracy of approximation. As a consequence, considering the low-temperature limit we infer a new non-asymptotic guarantee of convergence of the penalized gradient flow for the optimization problem.

## Poster 6 - Risk bounds for aggregated shallow neural networks using Gaussian priors

**Speaker:** Laura TINSI - [laura.tinsi@ensae.fr](mailto:laura.tinsi@ensae.fr)

**Abstract:** Analyzing statistical properties of neural networks is a central topic in statistics and machine learning.

However, most results in the literature focus on the properties of the neural network minimizing the training error. The goal of this paper is to consider aggregated neural networks using a Gaussian prior. The departure point of our approach is an arbitrary aggregate satisfying the PAC-Bayesian inequality. The main contribution is a precise non asymptotic assessment of the estimation error appearing in the PAC-Bayes bound. Our analysis is sharp enough to lead to minimax rates of estimation over Sobolev smoothness classes.

### Poster 7 - Transferable Deep Metric Learning for clustering

**Speaker:** Mohamed ALAMI CHEHBOUNE - [mohamed.alami-chehboune@polytechnique.edu](mailto:mohamed.alami-chehboune@polytechnique.edu)

**Abstract:** Due to the curse of dimensionality, clustering in high dimension spaces remains a hard task mainly because distance-based algorithms like k-means are no longer tractable or effective. Moreover, the choice of the metric is crucial as it is highly dependent on the dataset characteristics; Euclidean and other standard distance metrics may not be appropriate. We propose a framework for learning a transferable metric. Using a graph auto-encoder, we show that it is possible to build dataset independent features characterising the geometric properties of a given clustering. These features are used to train a critic that serves as a metric which measures the quality of a clustering. We learn and test the metric on several datasets of variable complexity (synthetic, MNIST, SVHN, omniglot) and achieve close to state of the art results while using only a fraction of these datasets and shallow networks. We show that the learned metric is transferable from a dataset to another even when changing domain or task.

### Poster 8 - Concentration Inequalities for Two-Sample Rank Processes with Application to Bipartite Ranking

**Speaker:** Myrto LIMNIOS - [myrto.limnios@gmail.com](mailto:myrto.limnios@gmail.com)

**Abstract:** The ROC curve is the gold standard for measuring the performance of a test/scoring statistic regarding its capacity to discriminate between two statistical populations in a wide variety of applications, ranging from anomaly detection in signal processing to information retrieval, through medical diagnosis. Most practical performance measures used in scoring/ranking applications such as the AUC, the local AUC, the p-norm push, the DCG and others, can be viewed as summaries of the ROC curve. In this paper, the fact that most of these empirical criteria can be expressed as two-sample linear rank statistics is highlighted and concentration inequalities for collections of such random variables, referred to as two-sample rank processes here, are proved, when indexed by VC classes of scoring functions. Based on these non asymptotic bounds, the generalization capacity of empirical maximizers of a wide class of ranking performance criteria is next investigated from a theoretical perspective. It is also supported by empirical evidence through convincing numerical experiments.

### Poster 9 - Driving behavior Analysis from multi-modal data

**Speaker:** Zhegong SHANGGUAN - [zhegong.shangguan@ensta-paris.fr](mailto:zhegong.shangguan@ensta-paris.fr)

**Abstract:** Where Are You Looking At: The Cognition and Attention Analysis in Driving Behavior To prevent and mitigate traffic accidents caused by drivers, we try to identify the driver's style, behavior variance, and the relationship with the user's personality traits from time-series and image data. This poster introduced our ongoing research in multi-modal driving behavior data collection, psychological questionnaires design, and data processing methods.

### Poster 10 - How to assess the environmental impact of AI? This is the challenge of Sustainable AI Project

**Speaker:** Maya GUILLAUMONT - [maya.guillaumont@capgemini.com](mailto:maya.guillaumont@capgemini.com)

**Abstract:** In the actual context of a growing demand for eco-responsible IT, the issue of assessing the environmental impact of Artificial Intelligence has become a critical topic. To take an active part in this subject and contribute to the spread of good practices, the R&I department has started in 2020 the Sustainable AI (SusAI) project.

Indeed, latter advances in the development of AI algorithms and hardware have required more and more computational efforts and resources going along with a significant increase of the energetic and environmental costs. As a telling example, the training of some well-known Machine Learning algorithms is five times more polluting, in terms of CO2 emissions, than a car during its whole lifetime. Within this context, the main objective of SusAI is to evaluate the carbon footprint over the entire life cycle of several AI algorithms, from hardware devices to software such as framework and supporting libraries. Going through a Life Cycle Analysis (LCA) methodology, one of the main challenges rely on developing models for assessing environmental impact of AI from data acquisition to the end of model production and passing by algorithm development and training. In addition, the objective of SusAI is to provide valuable guidelines for the design and use of an eco-responsible and environmentally friendly AI.

### Poster 11 - Speech recognition under constraint

**Speaker:** Olivier MATZ - [olivier.matz@capgemini.com](mailto:olivier.matz@capgemini.com)

**Abstract:** Today, speech recognition technologies are mature enough to be integrated into marketable solutions. However, these are owned by large groups with access to significant resources, both from a computational and data point of view. In particular, the development of a voice recognition algorithm requires several tens of thousands of hours of transcribed recording to achieve the performance of solutions currently on the market. In addition, if voice recognition solutions are widely used in our daily lives, they require significant computing resources and are not deployed on device but via cloud computing services. In most cases, speech recognition offers do not meet the needs of industry because (i) they do not guarantee data confidentiality i.e., manufacturers refuse to send their private and sensitive data to the cloud, (ii) the cost of use is important and (iii) they are not adapted to the specific business vocabulary of the manufacturer. In addition, the state-of-the-art algorithms proposed in the literature are evaluated on datasets that are not representative of the real world (audiobook, recording in a silent environment, professional audio recording equipment). For example, our preliminary work on the subject showed that the precision of these state-of-the-art algorithms was not acceptable on real world data or in industrial conditions: word error rate greater than 40% and up to 90%. In this context, the development of voice recognition solutions deployed locally, requiring little data labeled for training and robust to noisy environments represents a real challenge in the field of industry.

To solve these challenges, our work focus on the development of voice recognition algorithm focuses on 3 axes:

- Develop a Speech to Text approach adapted to a small labeled dataset by focusing work mainly on recent approaches of self-supervised learning
- Reduce the complexity of algorithm for on-device deployment to overcome constraints related to cloud computing (cost, GDPR and data privacy compliant, ...)
- Robust the Speech to Text algorithms under real conditions (noise, reverberation, disturbance, ...)

This work was granted access to the HPC resources of IDRIS under the allocation 2021-AD011012570 made by GENCI.

References:

- [1] D. Amodei et al., 'Deep Speech 2: End-to-End Speech Recognition in English and Mandarin', 2015.
- [2] Q. Xu et al., 'Self-Training and Pre-Training Are Complementary for Speech Recognition', 2020.
- [3] A. Baeovski et al. 'wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations', 2020.

- [4] A. Défossez et al. 'Demucs: Deep Extractor for Music Sources with Extra Unlabeled Data Remixed', 2019.
- [5] A. Défossez, et al. 'Real Time Speech Enhancement in the Waveform Domain', 2020.