#Standards4AI

‘Putting Science Into Standards’ workshop

Welcome!
We will start soon

DATA QUALITY AND BIAS EXAMINATION AND MITIGATION IN AI

8 June, 15:45-17:15

Panel discussion

DATA QUALITY AND BIAS EXAMINATION AND MITIGATION IN AI

Roundtable speakers

Rasmus ADLER
Fraunhofer IESE

Francisco HERRERA
Granada University

David REICHEL
European Union Agency for Fundamental Rights

Fred MORSTATTER
Information Sciences Institute

Rapporteurs: Maurizio Salvi and Alexandra Balahur (JRC)

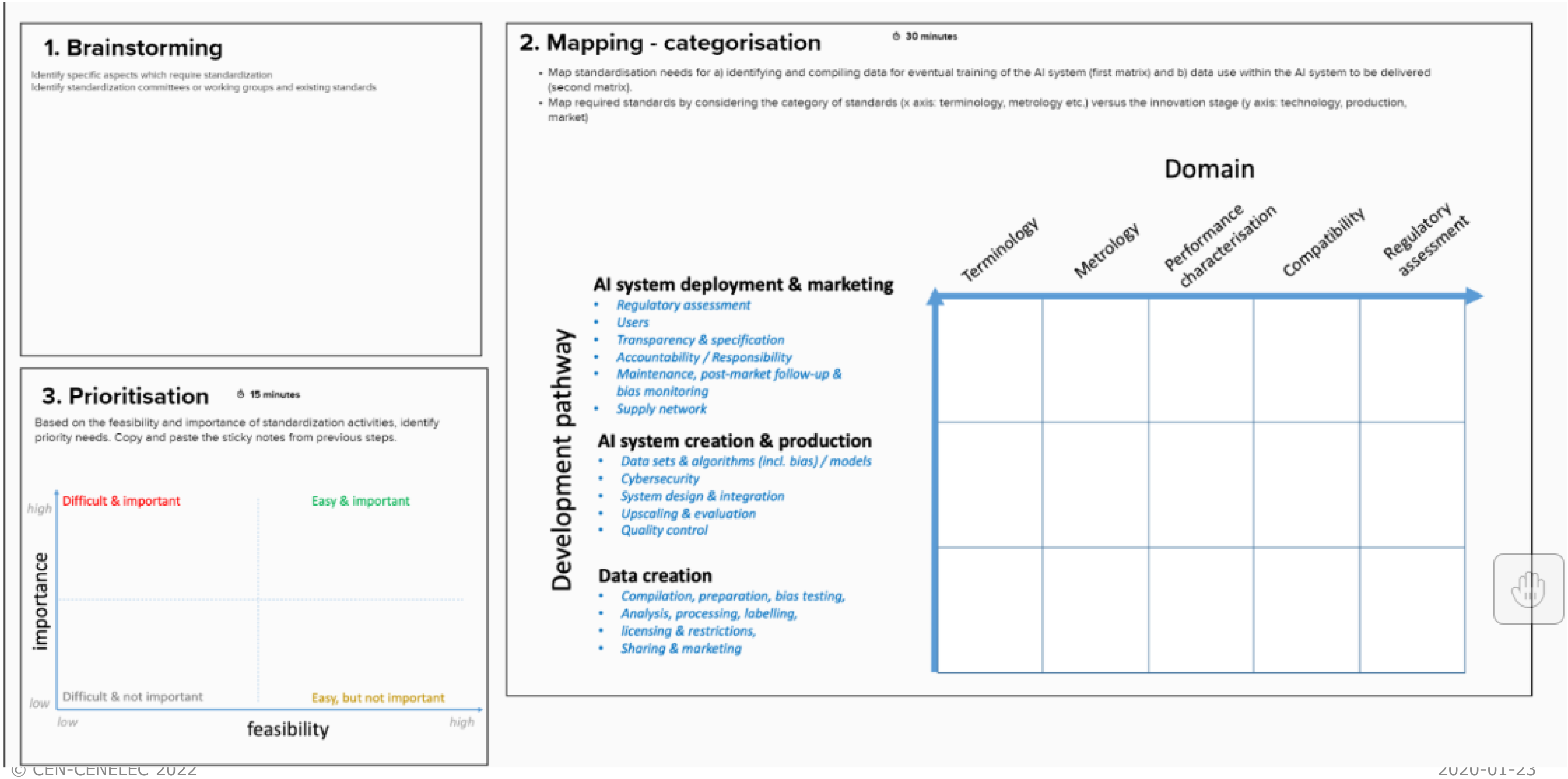
Audience interaction

slido.com

#Standards4AI

- Select the **Bias examination & mitigation** room on Slido
- Zoom chat - only technical questions to host
- Camera and audio OFF

Session structure



#Standards4AI

‘Putting Science Into Standards’ workshop

Rasmus ADLER

Fraunhofer IESE

Professional background

- Your background, linked to the topic of AI, particularly topic of this panel
  - Using AI in safety-critical applications / systems
  - Align research agenda of safety department and data science department with respect to the topic “safeAI”
  - Research question. ”Which data quality is sufficient when it comes to AI in safety-critical applications”?
- Line of work, type of projects
  - projects to give political recommendations; e.g. ExamAI https://testing-ai.gi.de/english
  - research collaboration projects like LOPAAS https://www.iese.fraunhofer.de/en/media/press/pm-2021-10-18-paradigmenwechsel-se.html
  - industry projects wrt. to topics various topics; Prominent topic is AI-based perception in automated driving (Hitachi, Bosch,…)

Challenges Faced & Solutions

- Related to data, bias in data, inclusiveness in AI, trustworthy AI in the topic of this panel
  - How to assure that the data for AI is sufficient wrt. usage of AI in safety-critical application?
  - When is the point reached where more and better data has no longer a strong impact on the “likelihood” of safety-critical AI failure modes?
  - How to assure that this point is reached and sufficient?
- What are some solutions you found to these challenges?
  - Use assurance cases and a way to structure the argument wrt. data for AI
    - **Integrating Testing and Operation-related Quantitative Evidences in Assurance Cases to Argue Safety of Data-Driven AI/ML Components**
    - Increasing Trust in Data-Driven Model Validation, **Safe Traffic Sign Recognition through Data Augmentation for Autonomous Vehicles Software**,..
  - Taxonomy for building assurance cases (e.g. based on **Towards a Common Testing Terminology for Software Engineering and Data Science Experts**)
- What initiatives/recommendations/etc. do you use in your work related to the issues of bias, inclusiveness in data and AI and generally trustworthiness of AI related to the topic of this panel
  - Make which claims can be established based on the measures taken to deal with bias etc. : e.g. claim “the risks due bias are mitigated as much as reasonably practicable”
  - Monitor state-of-the-art and -practice in this dynamic field to support the claim that “we applied best practices”

Way Forward, Next Steps

- What do you think is missing in the current regulatory and standardisation landscape with regard to the topic of this panel?
  - Regulation: The objective that should be achieved is missing. (e.g. demonstrate that best practices is applied to deal with bias)
  - Standards: Goal-based standards showing how to achieve the regulatory objectives
- What do you think are next steps in this area?
  - A standard for building an assurance case for AI.
    - Similar like ISO/AWI PAS 8800 but sector-independent and possibly not only for safety objectives but also for other objectives
- What do you think the focus should be for the short and long term?
  - Transition from prescriptive regulation and standardization to more goal-based regulation and standardization using assurance cases

#Standards4AI

‘Putting Science Into Standards’ workshop

Thank you!

#Standards4AI

‘Putting Science Into Standards’ workshop

FRANCISCO HERRERA,

GRANADA UNIVERSITY, SPAIN

herrera@decsai.ugr.es

Professional background

- Professor in the Department of Computer Science and Artificial Intelligence at the University of Granada, Spain.
- Director of the Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), Spain.
- Active projects:

I. Trust-ReDaS: Trustworthy and Responsible Data Science: Applications, Complex and Smart Data, Advanced Machine Learning

II. GO IMAI: Wood Identification by Artificial Intelligence and Mobile Device

III. CEPAI: Methodologies for improving Data Quality, Fairness and Privacy in Artificial Intelligence.

Challenges Faced & Solutions

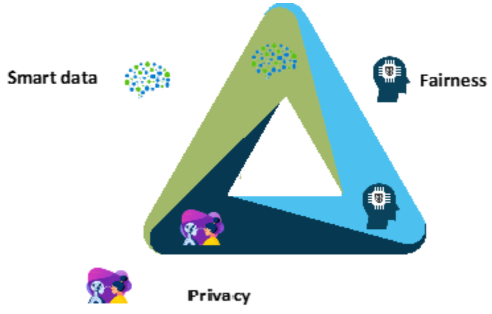
- Challenge 1: ***Learning from a data centric point of view***

- Classical approach "Model Centric view": Collect what data you can, and develop a model good enough to deal with the noise in the data.

- A new vision: The strengthening smart data to get quality data is the foundation for good artificial intelligence approaches from a data centric point of view.

Solution:

- To improve the data while hold the code/algorithm fixed is a necessary action to from a practical point of view.
- Consolidating smart data requires a data preprocessing analysis to adapt the data iteratively to fulfill the input demands of each learning algorithm.



Challenges Faced & Solutions

▶Challenge 2: ***To tackle the discrimination of AI algorithms***

Solution:

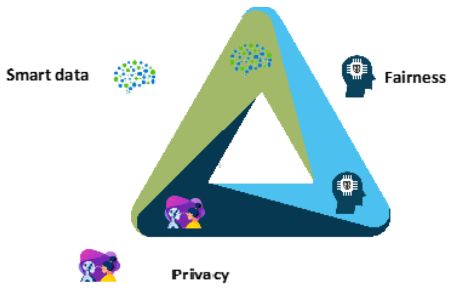
▶Transforming data and/or fitting models to obtain fairness in the decisions taken automatically

▶Challenge 3: ***Privacy***

▶Ensure privacy in the data and models obtained, avoiding the leakage of sensitive information

Solution:

▶Data privacy by design, without risk of being compromised.



Challenges Faced & Solutions

- (i) preprocessing of data (methodologies) to obtain Smart Data,
- (ii) avoiding bias in models through data fairness; (removing biased data and mitigating bias)
- (iii) robust learning to adversarial attacks that attempt to undermine the privacy of distributed data.

Way Forward, Next Steps

- ▶Artificial Intelligence Supervisory Agencies (coming)
- ▶Methodologies for moving from big data/noise data to quality data
- ▶Data silos (without human biases) that can be used together with small data sets of SMEs and that allow for a data-centric approach.

#Standards4AI

‘Putting Science Into Standards’ workshop

Thank you!

#Standards4AI

‘Putting Science Into Standards’ workshop

DAVID REICHEL

EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS

David.Reichel@fra.Europa.eu

Professional background

- European Union Agency for Fundamental Rights
  - Project manager, Data and Digital
  - Social science and data science background
- Current projects
  - Artificial intelligence and fundamental rights
  - Online content moderation and hate speech
- Relevant reports
  - Getting the future right – AI and fundamental rights (2020)
  - Data quality and AI – mitigating bias and error to protect fundamental rights
  - #BigData: Discrimination in data-supported decision making
  - Bias and algorithms (forthcoming in autumn 2022)

Challenges Faced

- Many examples of discriminatory use of AI
- One possible source of discrimination: data quality
  - Non-representative data
  - Low quality data
  - Erroneous data
  - Missing data
- Challenges to identify bias in relation to protected characteristics
  - Lack of awareness and assessments
  - Proxies and lack of data on protected characteristics
- Limits of mitigation

Solutions

- Awareness raising
- Legal requirements
- Fundamental rights impact assessments
  - Guiding questions on description of systems (documentation requirements) and assessments
  - including dataset descriptions
    - Lessons from statistical offices and social sciences

Way Forward, Next Steps

- Clear requirements for data and technology descriptions

- Engaging in policy making
  - **European Union (AIA)**
  - Council of Europe (CAI)
  - OECD (principles on AI)
  - UNESCO (recommendations)
  - …

- Additional research on concrete use cases

#Standards4AI

‘Putting Science Into Standards’ workshop

Thank you!

AI-Project@fra.europa.eu

#Standards4AI

‘Putting Science Into Standards’ workshop

FRED MORSTATTER,

USC INFORMATION SCIENCES INSTITUTE

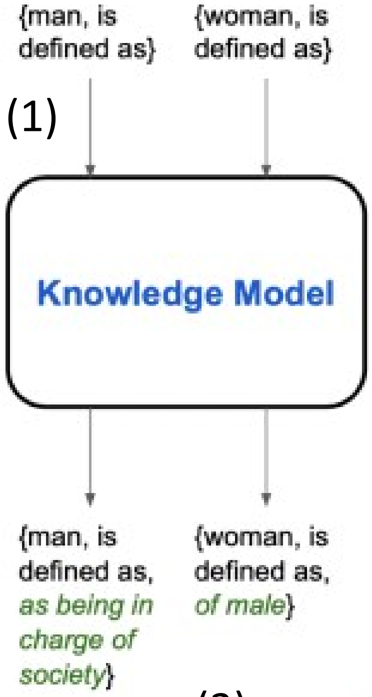
FRED@ISI.EDU

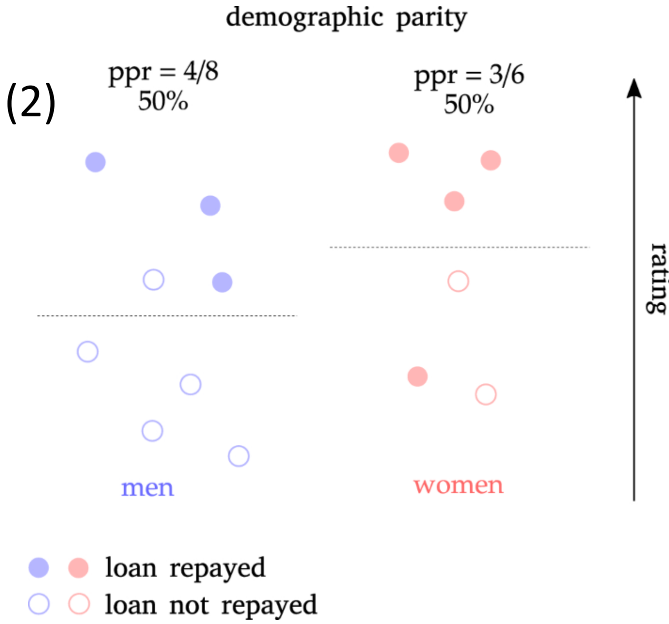
Professional background

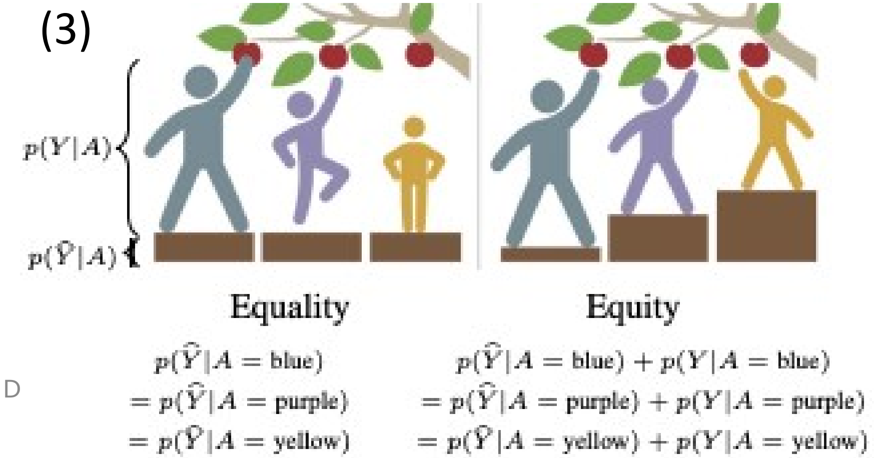
- Research Assistant Professor, University of Southern California
  Research Team Lead, USC Information Sciences Institute
- Active projects:
  - Fairness (definitions, data collection, applications)
  - Cultural modelling
  - Conversational Agents

Challenges Faced & Solutions

- Challenge 1: Biased Data
  - Data contains human biases.
  - Assessing, mitigating bias
  - Solutions: building robust models, removing biased data.
- Challenge 2: Choosing Fairness Metrics
  - Many fairness definitions, some incompatible, some missing.
  - Possible to game in certain contexts.
  - Solutions: Standardization in accepted contexts, multiple fairness metrics, working with stakeholders.



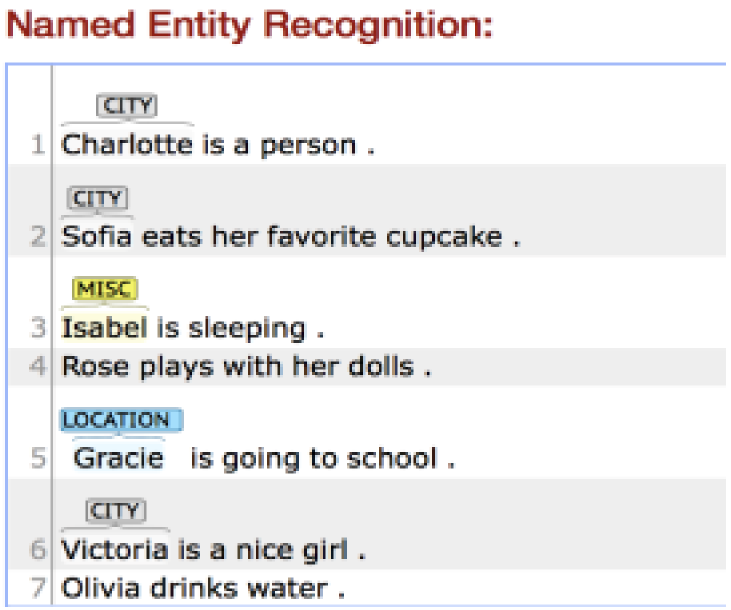


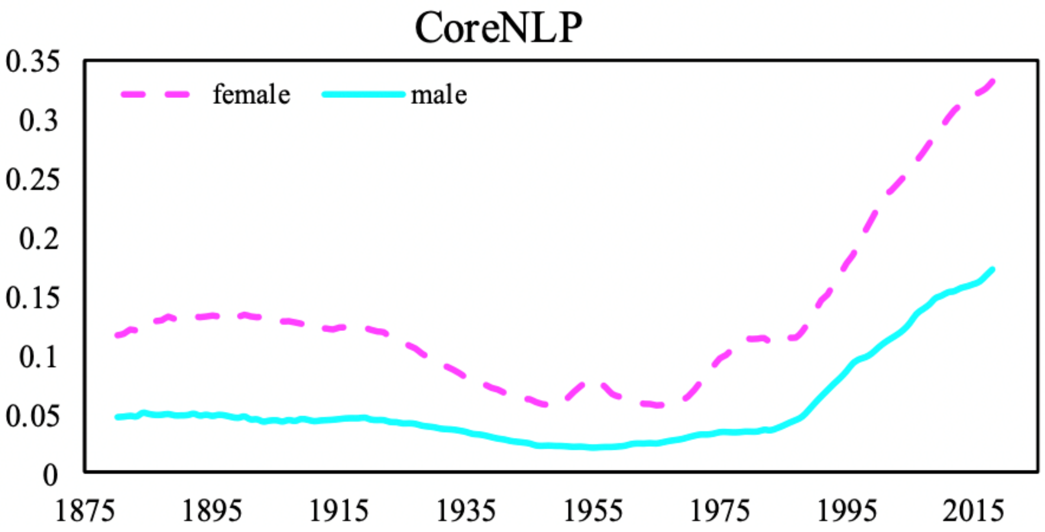


(1) Melotte, Sara, et al. "Where Does Bias in Common Sense Knowledge Models Come From." IEEE Internet Computing (2022).
(2) Castelnovo, Alessandro, et al. "A clarification of the nuances in the fairness metrics landscape." Scientific Reports 12.1 (2022): 1-21.
(3) Mehrabi, Ninareh, Yuzhong Huang, and Fred Morstatter. "Statistical Equity: A Fairness Classification Objective." arXiv preprint arXiv:2005.07293 (2020).

Way Forward, Next Steps

- Focus on open data, with selection and collection strategy
- Standardization of fairness metrics for high risk scenarios
- Auditing of algorithms
  - Direct predictors of high risk outcomes
  - “Upstream” predictors that are fed to models





#Standards4AI

‘Putting Science Into Standards’ workshop

Thank you!

Thank you for joining us today

See you tomorrow!