

# ILIdenT-Projektvorstellung





## Projektvorstellung

- ILIdenT = Intelligente Luftfahrttaugliche Ident-Technologien für die Supply Chain
- Durch die IFB Hamburg gefördertes Verbundprojekt zur Entwicklung verschiedener Identifikations- und Dokumentationslösungen für die Luftfahrtbranche
- Projektpartner:
- Use Cases, u.a. Erkennung verschiedener Bauteile (z.B. Fahrwerksbuchsen, Triebwerksrohre) & Defekte
- <a href="https://ilident.hamburg/">https://ilident.hamburg/</a>





























### Einordnung der Thematik

## Responsible/Trustworthy Al

### Forschung & Anwendung

- Ansatze, um z.B. Erklarbarkeit, Robustheit von KI zu erhöhen
- Vor- & Nachteile der Ansätze

### Rechtliche Anforderungen

• insb. AIA

### **Standards**

- Europäische Normen (CEN, Cenelec)
- EASA Fahrpläne & Konzeptpapiere

Was & wie? normativ

Was? verbindlich

Wie? *freiwillig* 







diskriminierungsfrei & nachhaltig (ökologisch, ökonomisch, sozial)

Ethics & Regulations



im Einklang mit europäischen Wertevorstellungen & geltendem Recht

Security & Robustness



konsistente Ausgaben trotz unerwarteter Eingaben

Governance



menschliche Kontrolle & klare Zuständigkeiten, Verantwortliche

Wechselwirkungen und Tradeoffs



Explainability



transparent & nachvollziehbar

Fairness & Sustainability



diskriminierungsfrei & nachhaltig (ökologisch, ökonomisch, sozial)

Ethics & Regulations



im Einklang mit europäischen Wertevorstellungen & geltendem Recht

Security & Robustness



konsistente Ausgaben trotz unerwarteter Eingaben

Governance





## Robustheit & Sicherheit

#### RAI

- Bandbreite/Variationen in Eingabedaten
- Generalisierbarkeit
- Drift-Phänomene im Betrieb
- KI-spezifische Sicherheitsaspekte

### AIA

- Artikel 9: Risikomanagementsystem
- Artikel 10: Daten und Daten-Governance
- Artikel 15: Genauigkeit,
  Robustheit und Cybersicherheit

- BM-1 bis BM-20
- u.a. zu Training & Testmethoden, (statistische) Validierungsmaße, systematische Erfassung von Risiken & Ableitung von Maßnahmen, Restrisiko



**Explainability** 



transparent & nachvollziehbar

Fairness & Sustainability



diskriminierungsfrei & nachhaltig (ökologisch, ökonomisch, sozial)

Ethics & Regulations



im Einklang mit europäischen Wertevorstellungen & geltendem Recht

Security & Robustness



konsistente Ausgaben trotz unerwarteter Eingaben

Governance





## Fairness & Nachhaltigkeit

#### RAI

- Definition & Ausdruck von Fairness in Form messbarer Größen
- Bias & Quellen (daten- & interpretationsgetrieben)
- Fairness in Anwendungsfällen ohne personenbezogene Date

### AIA

- Artikel 10: Daten und Daten-Governance
- Artikel 69: Verhaltenskodizes (freiwillig)

- BM-21 bis BM-30
- u.a. systematische Erfassung von Bias(quellen), Identifizierung von Untergruppen & Untersuchung des Modellverhaltens für Untergruppen, Sampling, fehlende/fehlerhafte Daten





**Explainability** 



Fairness & Sustainability



diskriminierungsfrei & nachhaltig (ökologisch, ökonomisch, sozial)

Ethics & Regulations



im Einklang mit europäischen Wertevorstellungen & geltendem Recht

Security & Robustness



konsistente Ausgaben trotz unerwarteter Eingaben

Governance





## Erklärbarkeit

#### RAI

- modell-inhärente vs. post-hoc Erklärbarkeit
- Vor- & Nachteile verschiedener XAI-Ansätze

### AIA

- Artikel 10: Daten und Daten-Governance
- Artikel 13: Transparenz und Bereitstellung von Informationen für die Nutzer
- Artikel 14: Menschliche Aufsicht

- BM-28 bis BM-42
- u.a. Datenvorverarbeitung, Feature Engineering & Selection, Auswahl eines geeigneten Modelltyps (Annahmen & Limitationen), XAI-Maßnahmen & Einschränkungen, Wiederholbarkeit, Form/Häufigkeit der Erklärung



Explainability



transparent & nachvollziehbar

Fairness & Sustainability



diskriminierungsfrei & nachhaltig (ökologisch, ökonomisch, sozial)

Ethics & Regulations



im Einklang mit europäischen Wertevorstellungen & geltendem Recht

Security & Robustness



konsistente Ausgaben trotz unerwarteter Eingaben

Governance





## Recht & Ethik

### RAI

- Umgang mit Daten
- Mensch-Maschine-Interaktion
- Abwägungen zw.
  verschiedenen Gütern
- Schutz der Grundrechte

### AIA

- Artikel 10: Daten und Daten-Governance
- Artikel 14: Menschliche Aufsicht

- BM-43 bis BM-47
- u.a. Datenschutz, Recht an Daten, systematische Erfassung der Risiken für Gesundheit, Sicherheit & Grundrechte



Explainability



transparent & nachvollziehbar

Fairness & Sustainability



diskriminierungsfrei & nachhaltig (ökologisch, ökonomisch, sozial)

Ethics & Regulations



im Einklang mit europäischen Wertevorstellungen & geltendem Recht

Security & Robustness



konsistente Ausgaben trotz unerwarteter Eingaben

Governance





### Governance

#### RAI

- Prozesse (Automatisierungsgrad/ human in the loop/ on the loop/ in command; Umgang mit Fehlerfällen)
- Verantwortlichkeiten
- Dokumentation

#### AIA

- Art. 9: Risikomanagementsystem
- Art. 10: Daten und Daten-Governance
- Art. 11: Technische Dokumentation
- Art. 12: Aufzeichnungspflichten
- Art. 13: Transparenz und Bereitstellung von Informationen für die Nutzer
- Art. 14: Menschliche Aufsicht
- Art. 17: Qualitätsmanagementsystem

- BM-48 bis BM-61
- u.a. geeignetes Maß an Automatisierung/menschlicher Aufsicht, Information & Schulung, Gewöhnungs-/Ermüdungseffekte, Gestaltung des User-Interfaces, Feedback-Mechanismus, Monitoring-Struktur, Ressourcenplan, Reaktion auf Fehlerfälle



### Standards & Compliance

- Die Agentur der Europäischen Union für Flugsicherheit (EASA) hat erste Dokumente veröffentlicht, die bereits angedachte Anleitungen & annehmbare Nachweisverfahren beinhalten
- Unterscheidung von drei KI-Leveln auf Basis des Automatisierungsgrads / Levels an menschlicher Aufsicht
- AI trustworthiness framework identifiziert vier Bausteine, wobei der erste Baustein als Schnittstelle zu den sieben HLEG-Kriterien für ethische KI dient:

### Level 1 AI: assistance to human

- Level 1A: Human augmentation
- Level 1B: Human cognitive assistance in decision-making and action selection

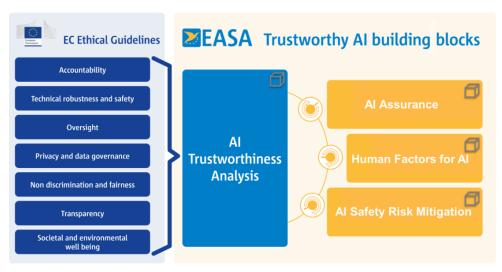
#### Level 2 AI: human-AI teaming

- Level 2A: Human and Al-based system cooperation
- Level 2B: Human and Al-based system collaboration

### Level 3 AI: advanced automation

- Level 3A: The AI-based system performs decisions and actions that are overridable by the human.
- Level 3B: The Al-based system performs non-overridable decisions and actions (e.g. to support safety upon loss of human oversight).

EASA AI Roadmap 2.0, Abbildung 4



EASA Concept Paper proposed issue 02, Abbildung 2



# Thank You!

Artificial Intelligence Center Hamburg (ARIC) e.V. Van-der-Smissen-Straße 9 22767 Hamburg

Office: +49 40 797 244 42 E-Mail: info@aric-hamburg.de

www.aric-hamburg.de

