2B or not 2B? The AI Challenge to Civil Aviation Human Factors

Barry Kirwan[1]

[1]EUROCONTROL, France

SUMMARY

Artificial Intelligence (AI) holds great promise for all industries in improving safe performance and efficiency, and civil aviation is no different. AI can potentially offer efficiency improvements to reduce delays and aviation's carbon footprint, while adding safety support inside the cockpit, enabling single pilot operations and the handling of drone operations in urban environments.

The European Union Aviation Safety Agency (EASA) has proposed six categories of Human-AI teaming, from machine learning support to fully autonomous AI. While AI support may in some cases be treated as 'just more automation', one category in particular, Collaborative AI (category 2B) considers the case of AI as an autonomous 'team-mate', able to take initiative, negotiate, re-prioritise and execute tasks. This category pushes the envelope when it comes to contemporary Human Factors evaluation of human work systems. The question arises, therefore, of whether Human Factors is sufficiently well equipped to support the evaluation and performance assurance of such new concepts of operation, or whether we need new techniques and even new frameworks for Human-AI teaming design and assessment. Four future Human-AI Teaming use cases are considered to help gauge where Human Factors remains fit-for-purpose, where it can be modified to be so, and where we may need entirely new techniques of performance assessment and assurance.

KEYWORDS

Artificial Intelligence, Human Factors, Aviation

Introduction

AI – Just more automation?

A definition of Artificial Intelligence (AI) is given by DeCanio (2016): "…*the broad suite of technologies that can match or surpass human capabilities, particularly those involving cognition*."

Generally, AI is about automating or supporting tasks such as problem solving and decision-making that are usually carried out by humans. AI can be seen as adding an additional enhanced (and non-deterministic) layer of support to the human, on top of 'traditional' automation. Categories found in AI are typically: learning, perception, reasoning, communication and knowledge representation. Applications include expert systems, machine learning, robotics, natural language processing, machine vision and speech recognition. Examples of early AI prototypes, products and services already exist in European aviation (European Commission, 2022), from automatic speech recognition and passenger support, to optimising safe and expeditious air traffic flow in normal and hazardous weather conditions. Some (e.g., Kaliardos, 2023) argue that AI innovation, for all its benefits, is essentially 'just more automation' supporting the human operator. In this case, the current Human Factors 'toolkit', perhaps with some minor modification, should be sufficient to enable successful integration of new AI support tools into the operational system. But AI is a fastdeveloping domain, and so an early evaluation of our readiness to support advanced AI introduction into complex, safety-critical industries such as aviation is warranted.

Narrow, Generative, and General AI

Most AI systems today represent what is known as *Narrow AI*, namely systems and services focused on a specific domain such as aviation. Narrow AI either supports humans in their analysis, decision-making and other tasks, or in cases where the tasks are well-specified and relatively predictable, it can execute its functions without human intervention and with minimal supervision. So far, this still sounds like 'just more automation', although the outputs of Machine Learning, for example, can be surprising, and go well beyond human analytic and computational capabilities.

The Human Factors challenge arises when tomorrow's AI systems are considered (e.g. see Salmon et al, 2020). *Artificial General Intelligence* (AGI, or General AI) does not yet exist, but is predicted to emerge in the coming decades. AGI would be capable of thinking and acting independently, of setting its own goals, and its intelligence could grow very rapidly to eclipse that of human beings. AGI, initially in the realms of science fiction, has been spectacularly foreshadowed by the arrival of ChatGPT[1], a Large Language Model that can give the appearance of intelligence. ChatGPT breaks the previously held conception of Machine Learning systems as essentially dumb number crunchers because of its interactivity, i.e. its apparent ability to have a normal (naturalistic) conversation with a human being.

ChatGPT and its correlates are *Generative AI* systems rather than AGI, as they are essentially carrying out calculations on a vast scale, rather than 'thinking' as such; they are very advanced and widely connected computers that are 'running the numbers'. There is no real intelligence there, as evidenced by the sometimes-strange answers such systems deliver, as they have no common sense. Rather, they are trained by humans (known as 'supervised training') to help them discriminate between acceptable and unacceptable responses. Although Chat GPT and similar systems are unlikely to be used directly in safety critical aviation operations, because they rely on the internet which is neither specific enough on aviation operations nor fact-checked, Generative AI opens up the possibility of interactive AI, which in turn has fuelled the notion of Human AI Teaming.

The Aviation Regulator's View on Human-AI Teaming

EASA (the European Union Aviation Safety Agency) guidance on AI currently proposes six categories of future Human-AI partnerships (EASA, 2023), interpreted by the author as follows:

- Machine Learning support (1A), already existing today;
- Cognitive Assistant (1B), equivalent to advanced automation support;
- Cooperative Agent (2A), able to complete tasks as demanded by the operator;
- Collaborative Agent (2B), an autonomous agent that works with human colleagues, but which can take initiative and execute tasks, as well as being capable of negotiating with its human counterparts;
- AI Executive Agent (3A), where the AI is running the show, but there is human oversight, and the human can intervene; and
- Fully Autonomous AI (3B), where the human cannot intervene.

In a recent debate on the issue of Human-AI Teaming (HAT)[2], a critical threshold appeared to be category 2B, since this is significantly different from what we have today. There are no systems in aviation currently that autonomously share tasks with humans, can negotiate, make trade-offs,

---

1. [1] Wikipedia on ChatGPT (2022) https://en.wikipedia.org/wiki/ChatGPT

2. [2] https://www.eurocontrol.int/event/technical-interchange-meeting-tim-human-systems-integration [Day 2]

change priorities and execute tasks under their own initiative. The question is therefore whether 2B is viable, and if so, what it means for Human Factors.

Aviation Human AI Teaming Use Cases

The Horizon Europe project HAIKU (https://haikuproject.eu/) is exploring six futuristic HAT use cases – two cockpit, two air traffic, and two airport – with varying levels of AI autonomy, including 2A, 2B and 3A. Four of the use cases are more team-oriented in nature (the other two are more similar to machine learning support for an airport, and use of a chatbot by passengers), as outlined below:

1. UC1 – a cockpit AI to help a single pilot recover from a sudden event that induces ‘startle response’, and directs the pilot to instruments to focus on in order to resolve the emergency situation. This cognitive assistant is 1B in EASA’s system, the pilot remaining in charge throughout.
2. UC2 – a cockpit AI to help flight crew re-route an aircraft to a new airport destination due to deteriorating weather or airport closure, for example, taking into account a large number of factors (e.g. category of aircraft and runway length; fuel available and distance to airport; connections for passengers, etc.). The flight crew remain in charge, but communicate/negotiate with the AI to derive the optimal solution. This is 2A but could also become 2B.
3. UC3 – an AI that monitors and coordinates urban air traffic (drones and sky-taxis). The AI is an executive agent with a human overseer, and is actually handling most of the traffic, the human intervening only when necessary. This is category 3A.
4. UC4 – a digital assistant for remote tower operations, to alleviate the tower controller’s workload by carrying out repetitive tasks. The human monitors the situation and intervenes if there is a deviation from normal (e.g. a go-around situation, or an aircraft that fails to vacate the runway). This is a cooperative agent, category 2A.

These use cases are in relatively early design stages, but prototypes are being developed such that already midway through the HAIKU project several are the subject of simulations (cockpit and air traffic virtual tower) with licensed crews (pilots and controllers) and preliminay AI interface designs. These use cases therefore serve as useful ‘testbeds’ for Human Factors approaches.

Human Factors Assessment of Human-AI Systems

Existing Human Factors Frameworks

A framework for Human Factors Assurance used frequently in European Air Traffic Management is the SESAR Human Performance Assessment Process (SESAR-HPAP), which has four high level requirements areas: human limitations and capabilities, the human-machine interface, teamwork and communication, and transition from design into operation. Additionally, the recently launched HURID (Human Risk-Informed Design) platform (which also includes SESAR-HPAP) from the European SAFEMODE Project (https://safemodeproject.eu/EhuridIndex.aspx) serves as a methodological toolkit for current and future systems validation from a human performance and safety perspective. The toolkit is illustrated in Figure 1.

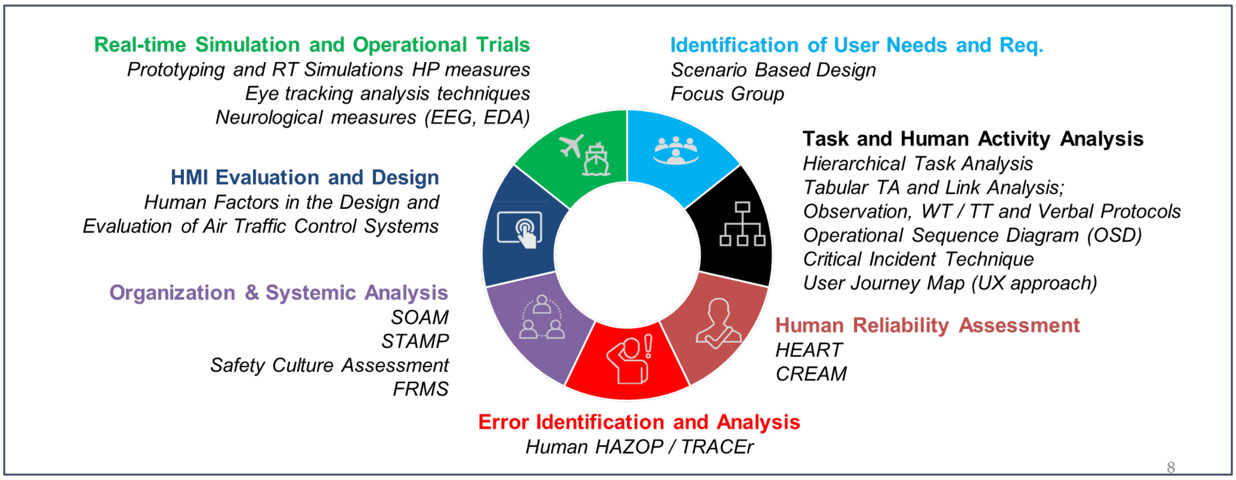

Figure 1: HURID Human Factors Toolkit (from the SAFEMODE Project)

Salmon et al (*op cit*) identified a number of Human Factors methods that could be applied to AGI systems, some of which overlap with those in Figure 2 (e.g. HTA and STAMP/STPA), and some others that do not appear in Figure 1 (e.g., EAST). Based on a review of such frameworks, a provisional Human-AI Teaming Human Factors Assurance Process was developed, shown in Figure 2, and is being applied to the four use cases.

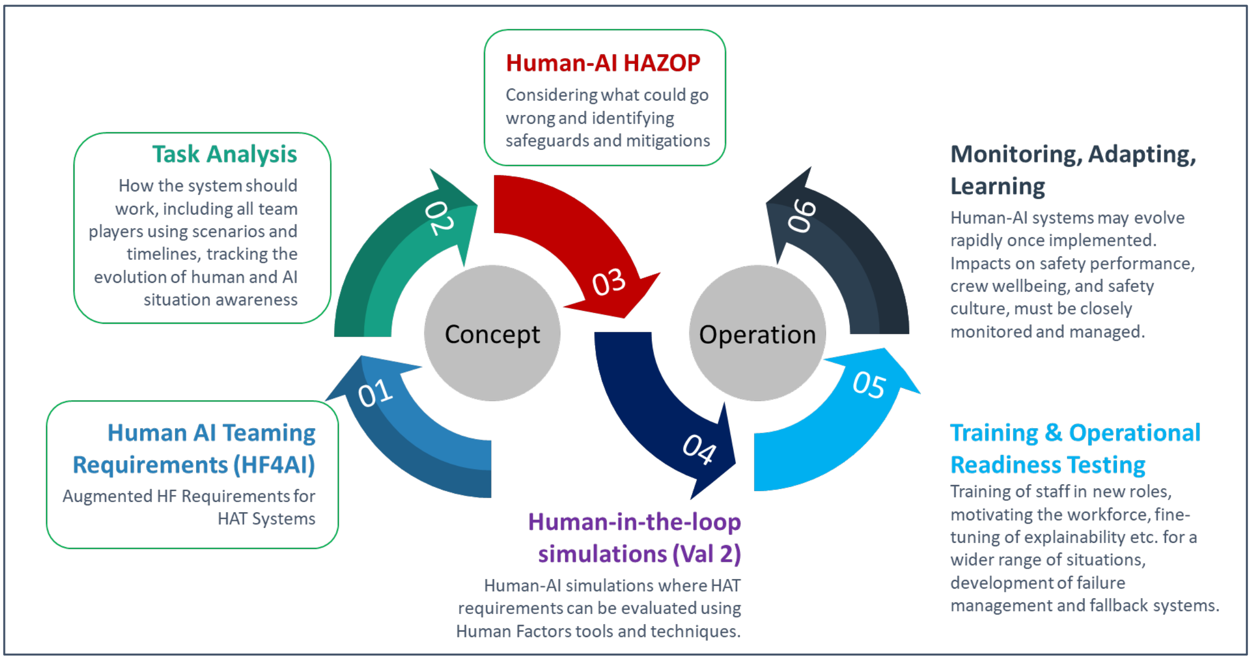

Figure 2: A Provisional Human Factors Assurance Process for Human-AI Teaming

Human Factors Requirements

The first element refers to Human Factors Requirements. The starting point used the SESAR HPAP requirements applied to new air traffic control systems, as well as existing regulatory guidance on cockpit design, for example EASA CS25.1302[3]. To these were added additional requirements questions purely focused on Human AI Teaming, based on an extensive literature review. In some cases, this meant adding items to traditional Human Factors Requirements areas such as Roles and

[3] https://www.easa.europa.eu/en/document-library/easy-access-rules/online-publications/easy-access-rules-large-aeroplanes-cs-25?page=41

Responsibilities, or to Equipment Design. In other cases, completely new requirements were needed, e.g. relating to explainability, shared situation awareness, and the skill-set required to determine when an AI is 'in error'. Example HAT questions given in Table 1 for Use Case 1.

Table 1: Sample Human Factors Requirement Questions and Responses (Use Case 1)

Requirement Related Question	Answer	Design Team Responce
	Roles & Responsibilities *(sample)*	
Are there any new roles, or suppressed roles?	Y	Single Pilot Operation concept, so one flight crew is no longer there.
What is the level of autonomy – is the human still in charge?	Y	The pilot remains in charge
Does this level of autonomy change dynamically? Who/what determines when it changes?	Y	If AI detects startle, AI begins support procedure for 20s, then directs attention to key cockpit instruments, then hands back control (T0 + Xs). Pilot can take control at any moment.
Does the human have responsibility for overriding the AI, or taking on its role if it fails?	Y	Pilot can override and cancel both startle support and directed instrument guidance
	Explainability *(sample)*	
Can the AI present information differently to resolve the misunderstanding?	N	Could add speech to CONOPS (some pilots already suggested it).
Can the human view both information or data that was used, as well as data that were ignored or discarded by the AI, e.g. anomalies/outliers?	N	Could be useful for post-event reconstruction / learning for both pilot and AI.
Does the timing of the explanation take into account the time-criticality of the situation, the needs of the user and the operational impact?	Y	UC1 is intentionally reactive rather than explanative, as dealing with time-pressed scenarios.
If events or inputs are outside the AI's operational boundaries, is the user alerted?	N	User not alerted but could have the information. Ask pilots at Val 2.

The requirements process takes around half a day to go through all the questions, depending on the maturity of the design description (called the 'Conops' or Concept of Operation). Some questions may have to be answered later, for example after a more realistic real-time simulation (referred to in the Table as 'Val 2'). Sometimes the questions lead to a new design feature which becomes a requirement, e.g., see the response to the first 'explainability' question in Table 1. The design team found it useful to go through all the (approximately 150) questions, as a way of checking their own design assurance. All answers will be reviewed and updated after Val 2 (Q4 2024).

Task Analysis: Adapted Operations Sequence Diagram

Task analysis is core to Human Factors, providing a blueprint of the human interactions with the system. For Human-AI Teaming, an essential focus is on whether the AI and the human crew are 'on the same page', or have a shared situation awareness. Of the various Task Analysis techniques available, Operations Sequence Diagrams (OSD – also called vertical timeline analysis) was selected, as it has a natural focus on interactions in time, in an event or task sequence. Figure 3 shows an extract from the adapted OSD for Use Case 4, wherein the AI (called ISA) detects an inbound aircraft catching up the one in front, and suggests and then executes a re-sequencing of aircraft. The AI is likely to detect this event before the controller will see it. The OSD makes it clear what the AI 'knows' (its contextual assessment) and what the controller knows (their situation awareness), as well as how the AI will signal the controller.

Time	Actual System State	Goal	Human1	Info sources (non-AI)	Operator believed system state	AI believed system state	AI solution
T0	Aircraft data suggest that an aircraft (BAW412) is going faster than expected and the sequence needs to be reshuffled	Resequencing of the aircraft's order to relieve ATCO's workload	ATCO is not aware of the possible changes at the moment	Data from aircraft	ATCO knows that the assistant will provide support if needed	ATCO is unaware of the aircraft state	
T0+10s	Resequencing begins and ISA signals the two involved aircraft (BAW412; KLM321)	Making the ATCO aware that a resequencing is happening	The ATCO becomes aware that the ISA is resequencing	Data from aircraft	ATCO knows that the assistant has been activated	ATCO is made aware of the possible changes	Divert attention to the two aircraft

AI HMI	H-AI Dialogue	Authority gradient	Decision / Action	HF Impact: trust, SA; startle / surprise; workload; engagement; competence
The AI HMI remains the same at the moment displaying the sequence	ISA's resequencing will start automatically. The ATCO supervises and makes decisions based on these suggestions.	ATCO is in control supervising the state of the system and the sequence		No negative impact on Human Factors in this scenario. The AI is simply monitoring the aircrafts and the ATCO is unaware.
The sequence's cell blinks in flashy magenta colour.On the radar, the speed data of BAW412 is highlighted in magenta.	ISA signals to the ATCO that a sequence is ongoing and ATCO waits.	ATCO is in control supervising the state of the system and the sequence	ATCO is processing the information and waits the end of the resequencing	ATCO gains situational awareness about the ongoing event

Figure 3: Operations Sequence Diagram Extract (Use Case 4)

The two Use Cases that have finalised their OSD found it useful as a way to deconstruct Human-AI interactions that can happen very quickly, providing a 'blueprint' of human-AI interactions and communication/explanatory modalities. The OSD is then used as the basis for the third step in the process, HAZOP (Hazard & Operability Study), which uses a set of guidewords to consider what could go wrong, and what mitigations to put in place to ensure / enhance resilience.

HAZOP

HAZOP is a structured brainstorming technique, involving a HAZOP leader, a note-taker, and experts from the design team, as well as operational experts (pilots; controllers) plus one or more HAIKU experts in safety and Human Factors. A missing partner so far has been the AI developer, but this is planned for the second wave of HAZOPs after completion of the validation experiments. An example of the output from Use Case 1 is shown in Table 2. The HAZOP approach has already found and offered mitigations for rectified potential vulnerabilities in the initial design and Conops.

Table 2: HAZOP Example Output (Use Case 1)

STEP	GUIDE-WORD	AGENT Human/AI	HMI	HAZARD	CAUSE(S)	CONSEQUENCE	EXISTING SAFEGUARDS	RECOMMENDATION 1. Operational 2. Validation 3. Concept
Situation Awareness Support activates	None	AI	Focus	Event outside Focus 'database'	Event not covered adequately by Flight Phase guidance based on pilot expertise	SA not supported	As-is today	Further (post-UC) assurance that flight phases cover all potential scenarios
	More	AI	Focus	Too much highlighting or too quick moving from one display to another	Design issue depending on scenario	Pilot confusion and loss of SA	Can disable SA support if not helping.	Consider prioritisation of key displays in the CONOPS.
	More	AI	Focus	Colour coding of highlighting used in Val 1	Confusion between priority and sequence (red is always top priority and can signify danger)	Use of Red for highlighting not preferred by pilots, as it sends a mixed message.	Disable Focus if confusing.	Consider avoiding red for highlighting.
	Other than	AI	Focus	Wrong highlighting on display	AI's assessment of what is happening, or the mode or flight phase is wrong	Safety risk	Disable Focus if pilot realises it is wrong.	Have one or two scenarios at end of Val 2 where the AI gives wrong advice, to see if pilots detect it.

Validation (Real-Time Simulation)

The fourth stage of the process involves real-time simulation and validation using Human Factors methods, including psychophysiological measures and eye-tracking, questionnaires (e.g. for workload and situation awareness) as well as 'soft' methods such as debriefs and focus groups. This will require subjective measurement of trust, utility of explainability features, and overall acceptability and desirability of the AI support system. The validation simulations (scheduled for late 2024) will also explore whether the human crews can detect erroneous AI outputs. The results of the validations will feed not only the design and Conops, but also training needs analysis and system deployment support, as well as identifying key performance parameters to monitor during initial operation (Steps 5 & 6 in the process in Figure 2).

Discussion

So far, the methods have mainly been applied to 1B and 2A systems, and are proving effective in identifying and 'fixing' potential human performance issues with the AI tools. It is likely that UC2 – which can reach category 2B – will challenge the methods due to the higher degree of autonomy, interactiveness and human-AI dialogue involved in this use case. But if the methods remain effective for this use case and for UC3 (which is category 3A), then it would appear that updating existing methods ('new wine in old bottles') may be sufficient for now. However, the analyses so far have already highlighted several areas where ideally new approaches, based on new research and possibly even new theory, are warranted.

For example, there is much talk of *shared situation awareness*. Yet AIs do not have awareness; rather they have a range of parameters that can be weighted in terms of their current relevance. Probably a better term is contextual assessment. The key aspect is to ensure that the human and AI are 'on the same page' in terms of what is going on and what the priorities are. But there are two elements of 'reflexivity' here: first, the AI may take as an input to its contextual assessment the human's interactions, inferring the human's SA. Second, the AI will likely be trained by humans, so there is a risk of being biased rather than being an independent 'second pair of [electronic] eyes'. This area may well need an entirely new model of compound, and dynamically interactive situation awareness.

Explainability is ripe for Human Factors research and innovation, rather than leaving it to the data scientists, especially as it is likely that users will want to customise their preferences in terms of explainability. HAIKU is exploring Construal Level Theory (McDermott and Folds, 2022) which affords a progressive system of explainability (from 'headline' to very detailed).

Validation measures will probably need to evolve, particularly with respect to trust, teamworking, collaborative working / interactivity, interface design and multi-media dialogue. Validation protocols themselves need to include scenarios where the AI 'gets it wrong', and also where the human gets it wrong and needs to be corrected by the AI. It may be that aviation can learn from other domains here, for example automotive, or military.

Pre-operational training and readiness testing will almost certainly need to involve more thorough understanding by the human operator of how AI systems work and can fail. Human supervisory training of AIs, and simulator training scenarios where humans and AI can 'build rapport' will need to occur prior to operational use. This area may require new models and new trainer knowledge and skill-sets, particularly as AIs may continually learn.

2B or not 2B?

The title of this paper alludes to the original Shakespearian existential question – in the context of AI referring to thinking, or even sentient AI (AGI), which does not yet exist. If and when AGI does arrive, it will be a true game-changer for society, and Human Factors will need to think again. Until then, there is no need to 'personify' or anthropomorphise AI, but there is a need to raise our game in supporting the human crews who will become users of advanced AI in safety critical domains such as aviation. Human Factors people need to work with Data Scientists, operational personnel (flight crew, air traffic controllers and airport personnel) and others on future Human-AI use cases, to evolve/develop techniques to bridge the gap likely to open up as AI technology develops. HAIKU and several other projects have made a start. It is time for Human Factors to invest in this area.

Funding

This publication is based on work performed in the HAIKU Project which has received funding from the European Union's Horizon Europe research and innovation programme, under Grant Agreement no 101075332. Any dissemination reflects the author's view only, and the European Commission is not responsible for any use that may be made of information it contains.

References

DeCanio, S. (2016) Robots and Humans – complements or substitutes? Journal of Macroeconomics, 49, 280-291.

EASA (2023) EASA Concept Paper: first usable guidance for level 1 & 2 machine learning applications. February. https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-20-published

European Commission (2022) CORDIS Results Pack on AI in air traffic management: A thematic collection of innovative EU-funded research results. October 2022. https://www.sesarju.eu/node/4254

Kaliardos, W. (2022) Enough fluff: Returning to Meaningful Perspectives on Automation. https://rosap.ntl.bts.gov/view/dot/64829

McDermott, T., Folds, D., 2022. Construal level theory in the design of informational systems. Frontiers in Physics, 10.

Salmon, P., Carden, A., and Hancock, P. (2020) Putting the humanity into inhuman systems: How human factors and ergonomics can be used to manage the risks associated with artificial general intelligence. Human Factors and Ergonomics in Manufacturing 31(4).

DOI: https://doi.org/10.1002/hfm.20883