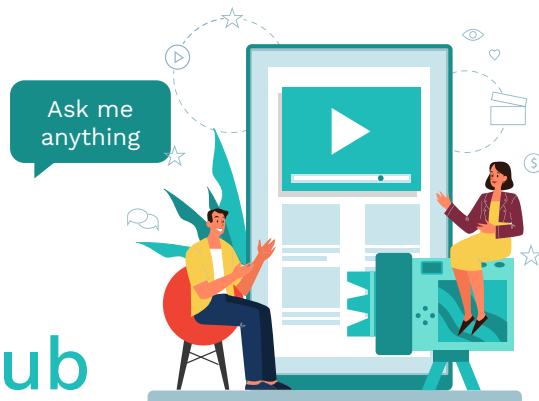


DiME

Virtual Journal Club



Large language model AI chatbots
require approval as medical devices

October 17, 2023 | 12 pm ET



Paul Wicks
Independent Consultant
Wicks Digital Health



Cécile van der Heijden
LLM / attorney-at-law
Axon Lawyers



Erik Vollebregt
Partner
Axon Lawyers



Stephen Gilbert
Professor of Medical Device Regulatory
Science
Dresden University of Technology



Tom Melvin
Associate Professor of Medical Device
Regulatory Affairs
**School of Medicine, Trinity College
Dublin**



Megan Coder
VP, Product & Policy
DiMe
Moderator

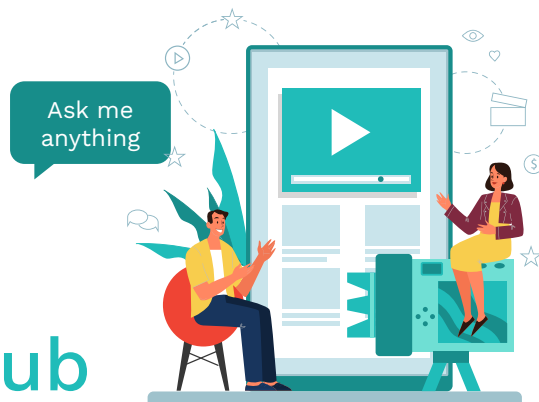
But first, housekeeping

- Please note today's session is being recorded
- To ask a question for discussion during Q&A, please:
 - Either 'raise your hand' in the participant window and come off mute when the moderator calls on you, or
 - Type your question into the chat box
- Slides and recording will be available after today's session
- Today's speakers represent their own views and opinions

Panelist Introductions

DiME

Virtual Journal Club



Large language model AI chatbots
require approval as medical devices

October 17, 2023 | 12 pm ET



Paul Wicks
Independent Consultant
Wicks Digital Health



Cécile van der Heijden
LLM / attorney-at-law
Axon Lawyers



Erik Vollebregt
Partner
Axon Lawyers



Stephen Gilbert
Professor of Medical Device Regulatory
Science
Dresden University of Technology



Tom Melvin
Associate Professor of Medical Device
Regulatory Affairs
**School of Medicine, Trinity College
Dublin**



Megan Coder
VP, Product & Policy
DiMe
Moderator

Article Introduction

“Laws for the Genie”

Article link: <https://t.co/ti7CvPRA84>

... the genie is out
of the bottle ...

*Images used under principle of fair use, without explicit permission for the purposes
of teaching, scholarship and research. Attribution: Disney (1992)*



... ban the genie -
put it back in the
bottle ...



... we can't put the
genie back in the
bottle ...

... rules needed ...



[comment](#) > [article](#)

Comment | [Published: 30 June 2023](#)

Large language model AI chatbots require approval as medical devices

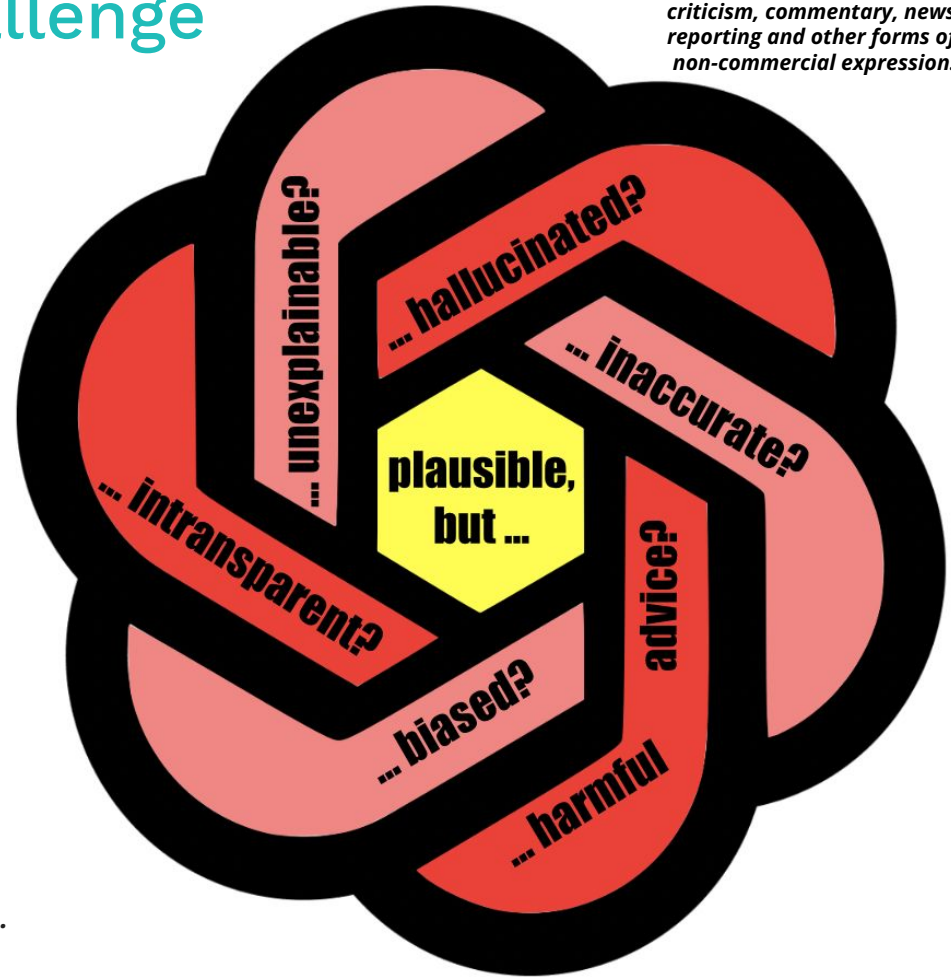
[Stephen Gilbert](#) ✉, [Hugh Harvey](#), [Tom Melvin](#), [Erik Vollebregt](#) & [Paul Wicks](#)

— [Show fewer authors](#)



The opportunity and challenge of shapeshifting LLMs

Use of the trademark of OpenAI in order to engage in criticism, commentary, news reporting and other forms of non-commercial expression.



Stephen Gilbert, Hugh Harvey, Tom Melvin, Erik Vollebregt & Paul Wicks.
Large language model AI chatbots require approval as medical devices. *Nat Med* (2023). <https://doi.org/10.1038/s41591-023-02412-6>

What is the problem ... ?


Table 1 | Challenges in the regulatory approval of large language models

| Challenge | Details |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Verification | Near-infinite range of inputs and outputs, including hallucinated outputs, make these models untestable |
| Provenance | No control over provenance when used as an underlying model on which a medical device is built |
| Changes | Not a fixed model, as the generative approaches and the manual and automated constraining of outputs (for example, to limit harmful advice) can be adapted on market |
| Usability | Near-infinite range of user experiences, depending on the input |
| Risks | No proven method to prevent harmful outputs |
| Surveillance | A near-infinite number of outputs make surveillance |

Context ... ?



Context ... ?

 **Dr. Gupta** CHAT


Dr.

Hello there! I'm Dr. Gupta.

Dr.

How can I help?

I have a headache



Dr.

I'm sorry to hear that you have a headache. I'd like to ask a few more questions to better understand your situation.

1. How long have you been experiencing the headache?

2. Can you describe the pain (e.g., dull, throbbing, sharp)?

3. Is the pain located on one side of the head or both?

4. Have you noticed any triggers for the headaches, such as certain foods or activities?

6 messages left. [Sign up for free](#) to send more messages. Message limits will reset on the first day of each month.

Send a message


SEND


Quality:


MEDIUM (FAST)

HIGH (SLOWER)

...

 **Dr. Gupta's Library**

 After you start your conversation, Dr. Gupta will collect research materials here.

 **Medical Information**

Provide your medical information for more personalized and informative suggestions.

IMPERIAL

METRIC (SI)

Age

yf

Weight

4

lb

Height


0

ft

0



in

Symptoms

headache 

Allergies

Medications

 **Vitals** 

Temperature

°F

Heart Rate

BPM

Respiratory Rate

breaths per minute

Oxygen Saturation

Waist Circumference

in

Hip Circumference


in

Systolic Blood Pressure

mmHg

Diastolic Blood Pressure

mmHg

 **Lab Test Results**

Albumin

g/dL

ALT

U/L

nitrogen (BUN) test reveals important about how well your kidneys are working. A is 7 to 20 mg/dL or 25 mg/dL.

BUN

mg/dL

Calcium

Creatinine

mg/dL

Glucose

mg/dL

HbA1c

mmol/mol

Potassium

mEq/L

Sodium

mEq/L

... LLMs,
decision
making, truth,
liability, and
the role of the
physician ...



Discussion

Table 1 | Challenges in the regulatory approval of large language models

| Challenge | Details |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Verification | Near-infinite range of inputs and outputs, including hallucinated outputs, make these models untestable |
| Provenance | No control over provenance when used as an underlying model on which a medical device is built |
| Changes | Not a fixed model, as the generative approaches and the manual and automated constraining of outputs (for example, to limit harmful advice) can be adapted on market |
| Usability | Near-infinite range of user experiences, depending on the input |
| Risks | No proven method to prevent harmful outputs |
| Surveillance | A near-infinite number of outputs make surveillance impossible |

Table 2 | Tasks and medical device status for large language models

| Task type | Use | Status in EU/UK ⁹ | Status in USA ¹⁰ |
|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Intake tasks | As a chatbot to assist patients to prepare for a telehealth consultation, such as suggestions on setting up equipment in the home ⁴ | Non-device | Non-device |
| | To assist with triaging non-critical patients presenting in the emergency department ⁴ | Medical device | Non-device, medical device or discretion on enforcement depending on details of function ¹⁰ |
| Core clinical tasks | Diagnostic decision support (DDS) ⁴ | Medical device | Medical device if providing specific advice and in time-critical scenarios |
| | Therapeutic-planning decision support (TDS) ⁴ | Medical device | Medical device if providing specific advice and in time-critical scenarios |
| | Counseling or talk therapy delivery ¹³ | Medical device | Medical device |
| Discharge tasks | Generation of clinical reports and related information ^{3,4} | Non-device or medical device depending on the intended purpose | Non-device if applicable requirements fulfilled ¹⁰ |
| | Generation of doctor's discharge letters ^{3,4} | Non-device or medical device depending on the intended purpose | Non-device if applicable requirements fulfilled ¹⁰ |
| | Generation of specific post-treatment patient information ^{3,4} | Non-device or medical device depending on the intended purpose | Non-device if applicable requirements fulfilled ¹⁰ |

Do you people just stand at the sidelines and talk of problems?



Table 3 | Steps for the approval of LLM-enabled medical devices

| Regulatory requirement and challenge | Possible approaches for developers |
|-------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A clearly defined intended purpose | Limit responses to user prompts to the stated scope of the device. Limit scope and severity to exclude emergency or critical use. |
| Determine the risk class according to national frameworks | Design to inform, not drive, medical decisions. |
| Define requirements | <ul style="list-style-type: none">• Implement performance benchmarks for narrow use cases. Stop or highly constrain on-market adaptivity, and document this.• Constrain the LLM to stop harmful advice and improve safety.• Control risks of data leakage and unconsented use of data to comply with data protection.• Provide accessible information to users on safe use. |
| Demonstrated analytical and clinical validity of output to inputs and use validatable knowledge base | <ul style="list-style-type: none">• Use only an LLM self-developed by the medical device manufacturer or an LLM developed by an external provider that is documented for use in a medical device.• Develop from authoritative medical sources.• Rigorously test, constrain, retest and document.• Link automated real-time fact checking approaches in feedback to the LLM.• Carry out comprehensive clinical trials to demonstrate the safety and efficacy of the LLM medical device, following regulations and reporting frameworks. |

Privacy and Protection

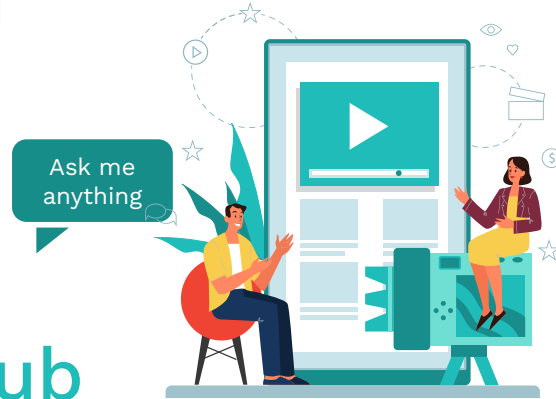


When a chatbot offers patients the ability to upload a complete patient medical record...

Attribution image: [David Goehring - Hippo Indigestion](#) ([source](#)) under cc 2.0 licence.

Questions

Virtual Journal club



Geeta Nayyar
Author, Advisor

Dead Wrong: Diagnosing and Treating
Healthcare's Misinformation Illness

November 14, 2023 11:00am ET



**Stay tuned for the 4th
anniversary webinar of our
Library of Digital Endpoints!**

Thursday, November 16

CROWDSOURCED LIBRARY
of Digital Endpoints





THANK YOU



@_DiMeSociety



[linkedin.com/company/dime-society](https://www.linkedin.com/company/dime-society)