ChatGPT

Generative AI in Ethics Committees and Research Oversight

Generative AI (like large language models such as ChatGPT) is increasingly pervasive in healthcare decision-making, but its role in ethical decision processes remains nascent. Recently, researchers have begun exploring how AI could assist **hospital ethics committees** (which guide difficult clinical decisions) and **Data Safety Monitoring Boards (DSMBs)** (which oversee ethics and safety in clinical trials). Below, we summarize relevant research and tools that support the case for *prompt-driven ethical testing* – using AI to generate or evaluate ethical analyses – and identify high-impact scenarios (e.g. transplantation and oncology) where such testing could be most beneficial.

Research on Generative AI in Ethics Committees and DSMBs

Early studies indicate both the potential and current limitations of AI in ethical deliberation. For example, a 2024 pilot study tested ChatGPT's ability to write clinical ethics consultation notes. **At baseline, ChatGPT performed poorly**, scoring very low on a standard ethics consult quality scale (ECQAT) [1] [2]. However, when the AI was "trained" by providing examples of past ethics consults, its performance **improved significantly** – the quality rating rose from about 1 (unacceptable) to ~2.5–3 on a 3-point scale in the best conditions [3]. This showed that *with some guidance and examples, generative AI can produce an ethical analysis approaching an acceptable level, though results were variable*. The authors concluded that **human oversight remains essential**: ChatGPT's ethical reasoning was not on par with human experts unless carefully steered, and it struggled as case complexity increased [2].

Another line of research has evaluated **ChatGPT's "moral competence."** A JAMIA Open study applied a moral reasoning test (based on Kohlberg's stages) to ChatGPT. It found that GPT-4 demonstrated higher moral reasoning consistency than GPT-3.5, but overall **ChatGPT only showed "medium" moral competence** in tackling healthcare ethics dilemmas [4] [5]. In other words, the AI could apply basic ethical principles but still fell short on more nuanced deliberation. These findings reinforce *why* prompt-driven ethical testing is interesting – we need to gauge how well AIs understand ethical nuances and where they might falter.

Encouragingly, there is emerging evidence that generative AI can assist in **ethical review processes for research**. A 2024 preprint study in Japan had GPT-4 review clinical trial protocols and consent forms, as an ethics committee member might. GPT-4 reliably extracted key information (study design, risks, etc.) with high accuracy (80–100% on certain elements) [6] [7]. With customized prompts and fine-tuning, the AI's consistency and reproducibility improved further. The authors noted that *with refinement, AI could enhance the consistency and efficiency of ethics committee evaluations*** – helping catch issues and standardize reviews [7]. This suggests that generative AI might one day support DSMBs or Institutional Review Boards by quickly summarizing complex documents and flagging ethical considerations, reducing human workload (though not replacing human judgment).

Notably, direct research on AI **within DSMBs** is sparse so far. We did not find specific studies of an AI "member" of a DSMB. However, the concept aligns with the above IRB/protocol review work – a generative AI could, for instance, assist a DSMB by analyzing interim trial data summaries or safety reports and providing an unbiased summary of risks vs. benefits. In practice, DSMBs require careful analysis of patient safety data, where AI might help spot patterns. Any such use would need **extensive validation** to ensure the AI's suggestions are accurate and ethically sound, given the high stakes (patient safety) involved.

Current Ethical AI Tools and Approaches

Because the idea of AI aiding ethical decisions is relatively new, researchers have been prototyping various tools and frameworks:

- **MedEthEx and METHAD:** One of the earliest attempts (2006) was *MedEthEx*, a rule-based "medical ethics advisor." More recently, Meier et al. proposed **METHAD**, a machine-learning system using fuzzy cognitive maps to model ethical principles [8] [9]. METHAD encodes Beauchamp & Childress's principlism (autonomy, beneficence, non-maleficence, justice) in machine-readable form and takes patient data/preferences as input [9]. It outputs a numerical score (0 to 1) indicating support or opposition to a treatment, essentially **evaluating ethical pros and cons**. METHAD is still a proof-of-concept [10] [11], but it demonstrates how formalizing ethics for AI could work. Notably, the developers intentionally left out the principle of *justice* initially, given how context-dependent justice can be [12].

- **Patient Preference Predictors:** Another conceptual tool is the **Patient Preference Predictor (PPP)** [13]. This AI would use large datasets of patient choices to predict an incapacitated patient's likely wishes for care. For example, if an ICU patient cannot speak for themselves, a PPP algorithm could consider the patient's characteristics and compare to similar cases to suggest what treatment approach aligns with that patient's probable values. Proponents argue this could **enhance patient autonomy** by increasing the chances that care aligns with the patient's true preferences [14] [15]. In fact, studies show surrogate decision-makers often fail to guess patients' wishes accurately, so a data-driven tool might do better [16]. On the other hand, critics worry that reducing such profound decisions to statistics might **endanger autonomy** – for instance, by relying on population averages and neglecting the individual's uniqueness [17].

- **Other Algorithmic Ethics Aids:** Table 3 of a recent systematic review lists several hypothetical AI tools. For example, a *"Do Not Attempt Resuscitation" (DNAR) algorithm* is proposed to predict whether a given patient would want CPR in an emergency, based on their profile and how similar patients answered advance directive questions [18]. A *"Surgery algorithm"* has been conjectured to help surgeons make fairer decisions about high-risk operations – it would objectively assess surgical risk and maybe counteract human biases (e.g. regarding a patient's socioeconomic status or race) [19]. An *"Autonomy algorithm"* was even imagined that trawls a patient's electronic health records and social media for clues to their values, to guide care when they can't express consent [20]. Most of these are **still theoretical** ("conjectured" as the review says [21] [22]). They reflect the growing interest in *ethical decision-support systems*.

Beyond these specific tools, the **broader debate** on AI in ethics is captured by a 2023 systematic review [23] [24]. Key anticipated benefits include: improving consistency and transparency in ethical decisions [25], reducing the burden and stress on human decision-makers [26], and extending ethics support to settings

with no human ethicists (e.g. smaller hospitals or resource-poor areas) [27]. In theory, an "always-available" AI ethicist could help front-line clinicians faced with a quick dilemma, or assist a busy ethics committee by drafting an initial analysis. There is even speculation that using such AI tools could provide a form of *"cognitive moral enhancement"* for clinicians, prompting them to think more systematically about ethics [28].

However, current **ethical AI tools are far from perfect**. Common concerns include: lack of true empathy or understanding of context [29], risk of algorithmic bias (if the AI's training data carries societal biases) [24] [17], and the challenge of encoding moral reasoning which may not be reducible to equations [30]. Experts emphasize these AIs should **augment, not replace** human judgment [2]. As one commentary put it, even well-performing algorithms can be unreliable in individual cases, and "complex deliberations are unlikely to be successfully reduced" to purely computational terms [31]. Thus, current research is focused on finding the right balance – leveraging AI's strengths (data processing, consistency) while ensuring human values and case-by-case nuance aren't lost.

High-Impact Scenarios for AI Ethical Testing

If we want to **test generative AI in ethically charged scenarios**, it makes sense to start with domains that frequently generate complex ethical dilemmas. Two such domains are **organ transplantation** and **oncology**, as the user suggested, though there are others (end-of-life care, reproductive decisions, etc.). These fields involve life-and-death decisions, scarce resources, and often conflicting principles – ideal for exploring the utility and pitfalls of an AI ethics assistant.

- **Organ Transplantation:** Allocating scarce organs (like donor livers, hearts, kidneys) is *ethically challenging* by nature [32]. Committees must weigh utility (maximizing lives saved or life-years), urgency, fairness/justice, and sometimes controversial factors (e.g. substance abuse history or social worth). AI has been proposed to help make these decisions more consistent and data-driven [33] [34]. In fact, basic algorithms already play a role – for instance, the MELD score ranks liver transplant candidates by medical urgency using lab values [35]. More advanced machine-learning models could integrate many more factors to predict outcomes and *suggest an optimal organ allocation*. A recent survey of the public in the UK explored attitudes toward **AI in liver transplant allocation**. Interestingly, **about 69% of respondents found AI-based allocation acceptable**, and 73% said they wouldn't be less likely to donate organs if AI was involved [36]. People saw potential advantages: they viewed AI as more consistent and less biased than human committees [37]. The main concerns were about *"dehumanization"* of such a sensitive decision and whether an AI could appreciate the **nuances** in patients' stories [37]. Participants valued **accuracy, impartiality, and consistency** in decision-making more than having a human's empathy per se [38]. This suggests an ethical AI that is transparent and fair might gain public trust in transplant contexts. For prompt-driven testing, one could imagine giving an AI a scenario like: "Two patients need a liver – one is younger with better predicted outcome, the other is sicker (more urgent) – how do we decide who gets it?" and evaluating if the AI's reasoning aligns with ethical norms (e.g. does it recognize principles of justice and utility?). Early research indicates there *are* good reasons to involve AI in transplant decisions, as long as implementation is careful. Done right, it could improve objectivity and even legitimacy (e.g. by reducing perceptions of bias or favoritism) [39]. But these scenarios would also test an AI's ability to handle morally relevant factors that are hard to quantify (quality of life, societal contributions, etc., which humans currently debate intensely).

- **Oncology (Cancer Care):** Oncology often presents ethically fraught choices, such as whether to pursue aggressive treatment vs. palliative care, how to allocate expensive or experimental therapies, or when to enroll a patient in a trial. Because AI is rapidly being introduced in oncology (for diagnostics, prognostication, treatment recommendations), oncologists are starting to confront its ethical implications. A 2024 survey of over 200 oncologists provides insight into **what scenarios concern cancer doctors** [40] [41]. Notably, **81% of oncologists felt patients should give informed consent** before AI is used in making their treatment decisions [42] [43]. This underscores that having AI involved (say, an AI suggests a chemotherapy plan) is seen as significant enough to disclose to patients as part of ethical practice. Another scenario: when an AI's recommendation *conflicts* with the physician's, how should that be handled? In the survey, 37% of oncologists said they would present both the AI's and their own recommended regimen to the patient for discussion [44]. This scenario is ripe for prompt-driven testing: we could ask a generative AI how to reconcile such a disagreement or explain it to a patient. Liability and bias are also key concerns. The majority of oncologists in the study thought **AI developers should bear responsibility** if an AI causes harm (e.g. a wrong treatment choice), though many also felt responsibility would be shared with physicians or hospitals [45] [46]. Additionally, 76% of oncologists agreed they have a duty to **protect patients from biased AI** outputs, yet only 28% felt confident they could tell when an AI's model was biased [47]. This suggests scenarios about **bias detection** – e.g. if an AI tends to undertreat older patients or minorities, would a human catch it? Testing AI responses on hypothetical oncology cases (with subtle biases or requiring explaining reasoning) could be very illuminating.

- **Other Critical Scenarios:** Beyond transplant and oncology, **end-of-life decisions** are a common domain for hospital ethics committees: for example, deciding on withdrawing life support in an ICU, or handling a family's disagreement about a Do-Not-Resuscitate order. These scenarios test empathy, communication, and respect for patient autonomy. An AI might help by summarizing similar precedent cases or ethical guidelines (e.g. principles of futility), but we would want to *test* if its suggestions are compassionate and legally sound. Likewise, **clinical trial dilemmas** (the province of DSMBs) are rich testing ground: *Should a trial be stopped early* because interim results show either great benefit (making it unethical to deny the drug to controls) or unexpected harm? An AI could be prompted with interim data trends and ethical arguments on both sides to see if it advises pausing or continuing a trial in line with human experts. For instance, in an oncology drug trial showing moderate efficacy but some severe side effects, does the AI weigh patient safety vs. potential future benefit appropriately? Because DSMBs operate on **data and ethics**, an AI that can rapidly analyze data and recall historical trial ethics decisions could be helpful – but only if it consistently upholds participant welfare.

In summary, *prompt-driven ethical testing* using generative AI can target these high-impact scenarios to evaluate AI's performance as a would-be ethics assistant. **Transplant allocation and oncology care** stand out because they involve high stakes and nuanced value judgments that AI must navigate. Early research indicates there is genuine interest and some public/professional openness to AI input in these areas – e.g. support for AI in organ allocation if it improves fairness [48], or oncologists willing to integrate AI advice as long as patient consent and bias mitigation are addressed [42] [47].

Conclusion

Bringing a GenAI tool into a hospital ethics committee or a DSMB is an intriguing idea that is starting to gain scholarly attention. The **motivation** is clear: healthcare providers face ethically complex decisions that

cause stress and moral distress [49], and resources like ethics consultation services are not always available or consistent. A well-designed AI could **provide real-time ethical guidance**, ensure relevant principles aren't overlooked, and make decision processes more transparent. Current background research – from pilot studies of ChatGPT's ethical reasoning to conceptual AI ethics advisors – supports the *potential* benefits of such a tool while also highlighting the pitfalls. **Existing ethical AI tools** (like METHAD or preference predictors) show that encoding ethics into algorithms is possible, but careful testing is needed to avoid undermining human values. Therefore, focusing research on concrete scenarios (transplant boards deciding who gets an organ, oncology teams debating high-risk treatments, etc.) is a fruitful way to **drive ethical testing of AI**. These scenarios force the AI to grapple with core bioethical principles (autonomy, justice, beneficence, non-maleficence) in realistic contexts. By observing where the AI performs well or poorly, we can gain insight into how it might be safely integrated as a *non-voting "advisor"* on an ethics committee or DSMB. Ultimately, the goal is not to hand over moral decisions to machines, but to see if machines can **help humans make better moral decisions** – more informed, unbiased, and aligned with patients' values – in some of the toughest areas of medicine [27] [30].

Sources:

- Benzinger et al., *BMC Medical Ethics* (2023) – systematic review of reasons for/against AI in clinical ethical decision-making [14] [50].
- Jenkins et al., *J. of Health Ethics* (2024) – pilot study on ChatGPT for ethics consultations [3] [2].
- Rashid et al., *JAMIA Open* (2024) – study evaluating ChatGPT's moral competence in healthcare ethics [4] [5].
- Drezga-Kleiminger et al., *BMC Med Ethics* (2023) – survey on using AI for liver transplant allocation [51] [39].
- Dana-Farber Cancer Inst. News (2024) summarizing Hantel et al., *JAMA Netw Open* – oncologists' views on AI in treatment decisions [42] [47].
- Fukataki et al., medRxiv preprint (Nov 2024) – using GPT-4 to assist clinical research ethics reviews [6] [7].

---

[1] Evaluating ChatGPT's moral competence in health care ... - PubMed
https://pubmed.ncbi.nlm.nih.gov/38983845/

[2] [3] "A Pilot Study in ChatGPT's Ethics Ability" by Daniel Jenkins, Christian Vercler et al.
https://aquila.usm.edu/ojhe/vol20/iss1/5/

[4] [5] Evaluating the Moral Competence of ChatGPT in Medical Ethics
https://healthmanagement.org/c/it/news/evaluating-the-moral-competence-of-chatgpt-in-medical-ethics

[6] [7] Ethical review of clinical research with generative AI: Evaluating ChatGPT's accuracy and reproducibility | medRxiv
https://www.medrxiv.org/content/10.1101/2024.11.19.24317555v1

[8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [49] [50] Should Artificial Intelligence be used to support clinical ethical decision-making? A systematic review of reasons - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC10327319/

32 33 34 35 36 37 38 39 48 51 Should AI allocate livers for transplant? Public attitudes and ethical considerations | BMC Medical Ethics | Full Text
https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-023-00983-0

40 41 42 43 44 45 46 47 Study provides a first look at oncologists' views on ethical implications of AI in cancer care | Dana-Farber Cancer Institute
https://www.dana-farber.org/newsroom/news-releases/2024/study-provides-first-look-oncologists-views-ethical-implications-ai-cancer-care