

We, the Healthcare Data Analytics team at Imaginary Health Analytics (IHA), in recognition of the impact our work has in driving healthcare initiatives and policy and the importance of securely and appropriately processing and utilizing protected health information (PHI) and personable identifiable information (PII), do hereby commit ourselves to the highest ethical and professional conduct and agree:

I. To uphold the highest standards of integrity, responsible behavior, and ethical conduct when performing analyses with All-Payer Claims Database (APCD) data, especially when PHI and PII is included.

1. to seek consent from all appropriate parties prior to project kick-off as part of an ethical project development process that ensures key stakeholders understand how and why certain data are being used;
2. to build and maintain client and stakeholder trust by only using data such as PHI or PII in appropriate situations that have been signed off on by clients or stakeholders in a statement of work, etc.;
3. to allow an individual full control of their data and maintain a system that allows individuals to easily opt-out of health insurance plans submitting their PHI and PII data to IHA;
4. to further build client trust by avoiding unlawful conduct and conflicts of interest, acknowledging and correcting project errors in a reasonable timeframe, and clearly outlining for colleagues and stakeholders the consequences that such conduct may have both professionally and in negatively impacting the direction of future healthcare policy and initiatives;

II. To treat all individuals fairly and with respect by ensuring appropriate steps have been taken to account for biases in the healthcare data.

5. to treat all individuals fairly, regardless of their race, age, gender, etc. by ensuring that minority groups are sufficiently accounted for in data sets used for analysis that will be used to drive healthcare policy changes nationwide;
6. to normalize data sets with the intention of eliminating factors and variables driven by the cumulative disadvantage often tied to minority populations;

III. To strive to ensure this code is upheld by all IHA employees.

7. to support colleagues in following this code of ethics, to strive to ensure the code is upheld, to not retaliate against individuals reporting a violation, and to ensure that all colleagues understand their ethical responsibility, regardless of their present role within the organization.

O/P/C Reference:

IEEE Code of Ethics. IEEE Advancing Technology for Humanity,
<https://www-ieee-org.proxy.lib.umich.edu/about/corporate/governance/p7-p8.html>.

Healthcare Data Analytics Pledge for the All-Payer Claims Database (APCD) Domain

Class Concepts: The Five Cs, Cumulative Disadvantage, Top Misconceptions about Data Science Ethics

The attached pledge was crafted for my company's work within the healthcare data analytics domain, specifically for All-Payer Claims Database (APCD) work. Within APCDs, PHI and PII is very prevalent and there is a comprehensive snapshot across time of enrollment and claims data for Medicare, Medicaid, and Commercial enrollees, resulting in these databases being powerful tools that drive healthcare initiatives and policy. As such, the goal of this adapted pledge would be to establish clear guidelines for a full project lifespan, from client consent and project design to implementation, while ensuring that all team members understand their ethical responsibility and the common biases to look out for when evaluating such a large but important dataset.

The IEEE Code of Ethics was the existing pledge that I adapted. It was chosen as it has three very strong "sections" (I-III) that dealt with ethical day-to-day behavior, the removal of biases, and the importance of all team members understanding their ethical responsibility, respectively. Therefore, I believed this pledge was a great starting place. From this initial pledge, which was unsurprisingly general as a global data science pledge, I updated language to tailor it to my organization. As part of this, I modified the Section I-III header language to align with this space, but the general themes remained unchanged. However, within each section, I updated all of the points to place greater emphasis on concepts like the Five Cs and cumulative disadvantage and questions around which team member(s) in an organization are responsible for addressing ethical issues when they arise.

The first section (I) of the updated pledge focused heavily on the Five Cs.¹ With this pledge, I hoped to make clear the importance of acquiring client consent and, in tandem, providing clarity around the data usage. Especially in the healthcare space where projects often involve PHI/PII and have high impact in driving policy, it is critical to ensure that clients' expectations are in alignment with the work being done, which will only build greater trust long term. Next, the second section was altered to focus more on data biases, particularly as it pertains to cumulative disadvantage.² This updated pledge puts a far greater emphasis on treating all races, genders, etc. equally and ensuring that appropriate steps are taken to account for minority populations, especially if initiatives or policies are intended for the society at-large. These actions are crucial, especially in the healthcare space, to make sure that policies are established that benefit all people, and this is often difficult to do, as there is always just less data available for these populations.³ Lastly, the third section handles a common misconception in data science that "ethics is someone else's problem."⁴ This section of the pledge was reworded to make clear that all colleagues in an organization have an obligation to think about the ethical ramifications of a particular project and, if there are ethical red flags emerging, to take action and ensure that the issue is addressed. With these updates in place, I believe this would be a strong pledge to put in place for my own organization and our own day-to-day work.

¹ Loukides, Mike, et al. "The Five Cs." *Ethics and Data Science*, O'Reilly Media, Inc., Sebastopol, CA, <https://learning-oreilly-com.proxy.lib.umich.edu/library/view/ethics-and-data/9781492043898/ch03.html#idm140316128118864>. Accessed 20 Nov. 2021.

² Sandvig, Christian. "Cumulative Disadvantage and Protected Classes." *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, <https://www.coursera.org/learn/siads503/lecture/5yC8Q/cumulative-disadvantage-and-protected-classes>.

³ Wallach, Hanna. "Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency." *Medium*, Medium, 23 Dec. 2014, <https://hannawallach.medium.com/big-data-machine-learning-and-the-social-sciences-927a8e20460d>.

⁴ Sandvig, Christian. "Middle Five Top Misconceptions About Data Science Ethics: The Countdown." *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, <https://www.coursera.org/learn/siads503/lecture/0GYZk/middle-five-top-misconceptions-about-data-science-ethics-the-countdown>.

References:

Loukides, Mike, et al. "The Five Cs." *Ethics and Data Science*, O'Reilly Media, Inc., Sebastopol, CA, <https://learning-oreilly-com.proxy.lib.umich.edu/library/view/ethics-and-data/9781492043898/ch03.html#idm140316128118864>. Accessed 20 Nov. 2021.

Sandvig, Christian. "Cumulative Disadvantage and Protected Classes." *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, <https://www.coursera.org/learn/siads503/lecture/5yC8Q/cumulative-disadvantage-and-protected-classes>.

Wallach, Hanna. "Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency." *Medium*, Medium, 23 Dec. 2014, <https://hannawallach.medium.com/big-data-machine-learning-and-the-social-sciences-927a8e20460d>.

Sandvig, Christian. "Middle Five Top Misconceptions About Data Science Ethics: The Countdown." *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, <https://www.coursera.org/learn/siads503/lecture/OGYZk/middle-five-top-misconceptions-about-data-science-ethics-the-countdown>.