WORKSHOP ON ETHICAL SOFTWARE ENGINEERING PRACTICES

Executive Summary

"What you cannot model, you cannot build" was the overarching theme of the **Workshop on Software Engineering Frameworks for Ethical Software** at ADCOM 2024. The pre-conference workshop, which focused on the ethical considerations behind the designing, developing, and deploying ethical software and systems, was led by IIIT Bangalore's Dean of Academics, Prof. Chandrashekar Ramanathan. Over the course of multiple interactive lectures with experts from academia and industry, the participants gained a better understanding of many of the real-world implications of the code they write, especially in a world moving towards increased automation. Consequently, this laid the foundation for discourse surrounding ethical and responsible software development.

Agenda

1. **Inaugural:** Introduction and Expectations (ethical considerations in software engineering) – Prof. (Dr) Chandrashekar Ramanathan, Dean (Academics), IIIT-Bangalore
2. **Foundation Talk 1:** Ethics by Design in Software Development – Mr. Prakash Narayanan, Counsel, IBM; Compliance Officer, IBM India and South Asia
3. **Foundation Talk 2:** Social Values and Principles in AI and Software Development – Prof. Frank Dignum, Wallenberg Chair in Socially Aware AI at Umeå University, Sweden; Director, Transdisciplinary AI for the Good of All (TAIGA)
4. **Foundation Talk 3:** Ethical Challenges of Bias in Data-Driven Applications – Mr. Vivek Karna, Co-founder, Zivy
5. **Foundation Talk 4:** Frameworks for Navigating Ethical Dilemmas – Ms. Srilakshmi Subramanian, Distinguished Member of Technical Staff, Wipro India
6. **Valedictory:** Closing remarks – Prof. (Dr) Chandrashekar Ramanathan, Dean (Academics) IIIT-Bangalore

Objectives

- Learning about the role of ethics in software engineering
- Understanding better the issues surrounding topics like bias, transparency, and data protection and applying this understanding to software engineering practices
- Designing, developing, and deploying the appropriate software engineering frameworks for specific ethical problems
- Making informed software engineering decisions that benefit people at large
- Understanding and incorporating the requisite international and national regulations along with the ethical requirements when developing software

Session Notes

Every IT application is a combination of hardware and software elements in order to find a solution to a given problem. These solutions are the most important aspect of any computer application, and generally fit into a few pre-determined categories like large-scale enterprise systems, smaller-scale consumer systems, and industry engineering systems, which, as the name suggests, are generally used exclusively by the engineering industry.

Engineering as a field requires engineers to be able to model first, and then build, immaterial of the surrounding circumstances. In software engineering, such modelling also establishes constraints and restrictions surrounding the solution the software problem. It is in establishing these constraints that ethics comes into play. The guardrail architecture for ethical software thus, consists of two main components—the model itself, and the ethics surrounding the modelling.

The most important question, thus, that remains to be asked is to explore what ethics means, both within and outside of the context of software engineering.

1. *Ethics by Design in Software Development*

In the first talk about ethics by design, IBM India and South Asia Legal Counsel and Compliance Officer Prakash Narayanan focused on the ethical aspects of AI with respect to tools and other software that software companies build for customers. Building AI and ML systems is not a fad anymore, rather a phenomenon that is here to stay. It is important to explore the idea of incorporating ethics by design because ultimately, systems learn from the inputs we provide—the outputs are dependent on the inputs provided by the people who design software systems. Considering this, it becomes necessary to explore if it is possible to make better business decisions while also focusing on creating ethical AI systems.

'Ethics' is branch of philosophy that studies moral principles. Concepts like 'morality', 'trust' and 'responsibility' play an important role in ethics in general. In software development, ethics can also include more specific and relevant aspects like transparency, accountability, privacy, and data protection. Ethics are bigger than laws, rules, and regulations, because even though they are generally based, at least to some extent, in ethics, they can be unethical. Further, the notion of ethics can change over time. The nuances of what ethics means to one person may differ from what it means to another—ethics can therefore also change from organisation to organisation, and even from society to society.

'Design' refers to creating or formulating an object, a system, or a process. Designing something often involves multiple steps like identifying the problem, researching, evaluating, prototyping, and improving on it, and finishing with the final product or project. Combining ethics with design, 'ethics by design' gives us a framework for integrating ethical principles into the design, development, and deployment of software, AI, or anything else.

Incorporating ethics into software and AI by design offers several benefits. For one, it can play an important role in solving for outcomes that may have significant moral or ethical hazards. For instance, it can help software developers design models without pre-conceived biases. Ethics by design can also help temper some customer expectations—generally, customers tend to prefer software with built-in guardrails to help them make better business decisions. Law always catches up with the ethical mores of society, so building software with the ethical considerations already in place can play a huge role in reducing future legal and regulatory risks for companies. Trustworthy and transparent products are always in demand, and so creating ethical products can also help companies distinguish themselves in the global marketplace. Subsequently, building ethical products and creating ethical technologies can become part of the organisation's culture and values. This can play a huge role in the larger context of the society as a whole, by creating positive social impact, but also can have a tangible impact on the company's ability to attract good talent, benefitting the company's bottom line as well.

Before designing ethical software (or anything else), it is important to formulate a plan to incorporate such ethical ideas and principles. A typical ethics by design plan consists of multiple steps, starting with specifying the objectives, and testing them against the organisation's ethics principles. Once the objectives have been specified and tested, the obvious next step is to specify the construction requirements, and detailing the if and how the developmental tools being used by the organisation will capture the ethical principles so specified. While doing so, whoever is building the software needs to assess the design specifications, resources, and requisite infrastructure against the organisation's Ethics Principles, and also crucially, look into conducting an Ethical Risk Assessment. It is also important to consider here what the outcomes will be, and whether they will be morally correct, or hazardous to the people. Typically, this is followed by High-Level Design functionality that supports the ethical requirements, like keeping internal data logs in case of data manipulation.

Perhaps the most important step in formulating an ethics by design plan is the collection and preparation of data. A good rule of thumb here is to always assume that data is biased, skewed, or incomplete, until it has been proven otherwise. Generally speaking, this is the step where a lot of issues come up, realistically speaking. Understanding and recognising potential issues in this area, as well as consulting with relevant stakeholders on the matter could prove very effective in making the data better. On the other hand, it is also important to recognise that there will always be some incompleteness in the data because ultimately, it is impossible to avoid all human error. The key lies in balancing good data collection with delivering the finished product to the market on time.

Following robust data collection with detailed design and development is an important next step. Developers will have to come up with ways to mitigate issues that come up during development, as well as reassess and iterate for new design details. Additional issues related to ethics may emerge whilst developing a product, raising questions about what may have been missed, and how this can be fixed. Finally, any product will have to be tested and evaluated before it is deployed. Creating a robust testing regime can be a very helpful tool in checking the ethical compliance of the final product. Altogether, actionizing ethics by design involves ensuring that the methodology used to create the product, the governance and other ethical principles involved, and the adoption of the final product are done whilst keeping ethical and moral mores in mind.

IBM, as an organisation, adheres to three principles of trust and transparency, as well as five pillars of trust to ensure that the software, AI, (and other) products are designed, developed, and deployed in a safe and ethical manner. Ethics by design is only part of IBM's overall organisational governance structure. The IBM ethics board has been responsible for commissioning and overseeing the company's ethics by design methodology. New projects of particular concern have to be approved by the ethics board, which makes these decisions based on both the principles and the pillars of trust.

IBM's Integrated Governance Program (IGP) works to ensure that all of the company's products and projects maintain compliance in the data protection and data governance spaces. This is iterative to the AI ethics aspects, and ensures that products are tested regularly against emerging regulations around the world. IBM Granite, for instance, has been built in compliance with IBM's ethics by design and IGP programs.

2. *AI for the Good of All? Social Values and Principles*

The second lecture by Prof. Frank Dignum the director of Transdisciplinary AI for the Good of All (TAIGA) at Sweden's Umeå University pivoted the conversation away from incorporating ethics into software design and development, instead focusing on the different social and cultural contexts that

define our everyday lives. Values and culture have become an important part of the AI conversation because the systems we build depend on our shared cultures and values, as well as our differences. The systems we build have to fit into both our lives, as well as our assumptions. Subsequently, the social realities and the contexts that we live in also determine the way we work and live, which also dictate the systems that we use. Putting AI in these social contexts, thus, becomes an important part of software development.

Through the lens of Dutch psychologist Geert Hofstede's cultural dimensions, Prof. Dignum explains the practical implementations of features in different software and systems across different cultural contexts. Per Hofstede, these six cultural dimensions—power distance, uncertainty avoidance, individualism-collectivism, masculinity, long-term orientation, and indulgence—determine how people behave in different countries. Consequently, what is 'good' and 'fair' is different in different countries. For instance, the United States and many countries in Wester Europe are more individualistic, with lesser power distance, while countries like Malaysia and the Philippines display greater collectivism with greater power distance. Typically, countries are scored across these dimensions from 0-100.

Countries differ in their behaviours across the dimensions, but these dimensions are, in many ways, universal. Naturally, this does not imply that every person in a particular country behaves in the same way based on these cultural dimensions, but rather that the people of a particular country share a culture. The importance of applying these dimensions lies in understanding that none of these dimensions are inherently 'good' or 'bad'. They do, however, provide a lot of context with regard to decision-making. These cultural dimensions are useful in determining values, which are also an underlying mechanism that drives human behaviour.

Similar to values, motives also act as primary drivers of human behaviour. Motives include factors such as achievement, affiliation, power, and avoidance, all of which determine our actions, as well as our interactions with other people. Human beings are also motivated by affordances and allowances. The end goal, generally speaking, is to get better as people, and improve our individual and collective behaviours.

Our motives and affordances tend to play an important role in how we function as a society. Being part of any society implies that the members of that society have things in common—they may typically be part of the same organisations, or they may observe the same practices, or follow the same conventions. A society also functions on the basis of social concepts like fairness, bias, equality, and inclusiveness. None of these words have a standard fixed definition, and none of this is based in any natural laws, but nevertheless, all of these aspects play an important role in defining the lives people build for themselves. Simply, society acts as the filter through which peoples values translate into action.

While discussions surrounding better understanding the society that we live in and its impact on our actions can seem more philosophical than technical, understanding the core similarities and differences between people, organisations, and societies is foundational in building ethical software that benefits the greater good. Consequently, balancing individual interests with the interests of society at large is key in taking the right actions at the right time. Ideas of fairness and equality can differ based on perspective, which also means that different perspectives on fairness can have different meanings and solutions for fairness for everyone.

Translating social values into the AI context involves first defining and understanding what both 'intelligence' and 'artificial intelligence' mean. Intelligence is the ability to perceive or infer information, and to retain it as knowledge to be applied towards adaptive behaviours within an environment or context. Artificial intelligence, as defined originally, refers to the research area that studies human intelligence through the design of computer systems that emulate (parts of) human intelligence. Based on this, it then becomes important to better understand what human intelligence means, rather than to build systems that that replace human intelligence. Human intelligence is inherently social—society is organised such that the things that get done in society get the society as a whole moving forward.

This is inherently different from the way that a traditional AI system works. Consider the supercomputer Deep Blue, for instance. Deep Blue was very good at playing chess, but it was only ever good at playing chess. That was the extent of its 'intelligence'. People on the other hand are typically more than one thing; a person may be good at playing chess, but she may also be a schoolteacher, a gardener, and a mother. In the context of a society, this also means that people are dependent on other people for their needs.

To add nuance, consider also, for example, the fact that we can determine whether or not people have cavities in their teeth using AI. This kind of pattern recognition is extremely important work, and in the recent past, has led to quicker and better cancer diagnoses for people. On the other hand, AI cannot answer more complex questions involving the rights and futures of people, like what the best way is to deal with refugees. Answering more nuanced question requires greater context, context beyond the powers of AI. This doesn't make AI bad, or useless, but rather limits its use to certain problems.

Even with pattern recognition AI systems, human intervention is important to be able to explain the final conclusions. In the context of using pattern recognition AI to detect cancer, recent studies have shown that the AI system and doctors looking at the scans can sometimes disagree on whether or not a particular image shows cancer. Here, an easy solution to the problem involves creating better dialogue with the radiologists, as well as the patients themselves to improve the AI system, leading to more human-centred AI.

Ultimately, in practice, the right explanation makes the difference in an AI system, as opposed to being the best, or having the best model. The explanations are more important than the results themselves, because the explanations help add context to the results. Moreover, not everyone can read and understand the more complex results generated by AI systems. In these cases, explanations by experts can make all the difference in improving the AI system.

Because of all the complexities that come with being human, and living in societies, being ethical whilst developing software is not limited to meeting some set of requirements. Rather, ethical issues have to be addressed in every step—through modelling, implementing, and ultimately, using the software.

An important part of this discussion also involves understanding where human emotions fall in this model. Emotions are perpendicular to the conversation around AI for social good, because emotions are reactions to situations, i.e., to the other societal drivers. What this means is also that emotions do not drive collective human behaviour, so while modelling for emotions is important and has its place, it is not relevant to the conversation about collective social good.

3. Ethical Challenges to Bias in Data-Driven Applications

Zivy co-founder Vivek Karna's talk focused more on the practical applications of and challenges to AI systems from a case-study perspective with real life examples such as financial inclusion and agentic AI. Financial inclusion, especially with respect to seeking funding and getting loans from financial institutions has changed with the introduction of algorithms into the systems that decide who gets a loan and who doesn't. Around the world, it is disproportionately affecting women, people of colour, low-income households, and the rural populations; or what the system calls 'high-risk' borrowers. Automatically, this leads to the question of who a high-risk borrower is, and whether banks and other financial institutions should lend to them.

To understand the arbitrary categorisation of people as 'high-risk borrowers', it is important first acknowledge that traditional credit score systems, which typically play an important role in making this kind of determinations, are inherently biased. Among other things, it only allows for users that are in the given system to qualify as borrowers, excluding an entire category of people who may be good at paying back loans, but just so don't happen to be in the system. As a result, typically marginalised groups are put at a disadvantage, based solely on historical data, and consequently, historical biases. The human impact of these decisions can be widespread and systemic, affecting already marginalised groups even more than before.

While there is always some risk that comes with lending to an entirely new category of people, especially one previously not part of the system, it is still possible to do so by avoiding pre-existing biases. One innovative solution, for instance, is to look at the cash-flow to the business (or other loan subject), as opposed to the existing credit situation (like credit scores) or lack thereof. Using non-traditional sources of data, thus, can provide a better, more accurate picture. Micro-lending companies, for instance, have developed alternative credit-scoring systems like 'karma scores', which are often based on alternative data sources outside of traditional credit-scores, such as repayment history or mobile phone usage.

However, on the flipside, it can also introduce new biases of its own. It is impossible to not have any bias at all—bias in some form or another is inherent, and not all biases are bad. In this context, however, it is important to have the self-awareness to understand biases in decision-making, whilst also being aware of the risks that come with making any decision, one way or another.

Other innovative ways to curb risks that around lending to high-risk borrowers could be to change the pre-existing monthly or quarterly repayment models into daily repayment models. Daily repayment could make the models more scalable, especially if such payments could be automated through a digital wallet. This could also be advantageous to the lenders, in that they would be able to find out sooner about any difficulties, financial or otherwise, with the business, if a particular day's payment is missed.

The period immediately after the COVID-19 pandemic saw a drastic rise in the number of predatory loan-making apps that targeted vulnerable people in India. This occurred because cloning an existing app, and then subsequently onboarding a lot of people onto the app is relatively easy. Large financial institutions have to report to multiple regulators and are regularly audited. Smaller lending apps, however, are subject to fewer regulations, meaning that it becomes very easy to misuse any data that is stored on these apps. App developers, software engineers, and other professionals involved in this market have to be more cognizant of what data is being stored and where. It is also necessary to consider what data is actually needed.

Data privacy is a key ethical consideration with respect to data collection, to ensure that personal data is both collected and stored securely, while protecting individuals' information from unauthorised access and abuse. Data security is different from data privacy, and robust data security will safeguard data against leaks, breaches, and other kinds of unauthorised access. It is equally important to obtain explicit, informed consent from individuals whose data is being collected. Informed consent means that the user must have an understanding of how and why their data will be used and stored. Finally, it is also important to recognise, and consequently mitigate biases in data collection, use, and analysis to ensure that people are treated fairly and without prejudice. With respect to practical implementation on a software level, ensuring that the software being used to build these systems have built-in data encryption, and creating better access controls can help ensure that data is stored securely.

The rise of agentic AI in the form of 'AI Assistants' could have serious impacts on the future of work. A typical AI assistant would work by reading messages, emails, and calendars, making phone calls, and performing other necessary tasks to increase productivity in a business setting, although there are other use cases for AI assistants in non-business settings. Balancing personalisation with responsible AI is very important when building an AI assistant.

A key aspect of creating a good AI assistant comes from the need for diverse and robust data sets to create good models. A good AI assistant is able to understand and respond to the unique business needs of the person seamlessly. However, this needs a lot of very specific, yet diverse data from unique use-cases. Personalising the model, therefore, requires good and accurate data, which has to be obtained and stored with some level of anonymisation. A good solution to this problem comes through zero-shot learning, and working with data the model has previously not seen or had any access to.

Outside the human impact of AI assistants, which may replace secretaries and other administrative assistants, there are a number of other ethical issues that have to be considered, including bias. Here, bias can be based in gender—for example, most AI assistants and other AI agents generally feature traditionally female-sounding voices, which could reinforce existing gender stereotypes. The bias may also be racial or socio-economic, generally due to limitations in the training data. Further, it is also important to ensure algorithmic fairness, transparency, and accountability when developing these systems, so that the creation of these AI agents remains as unbiased, fair, and trustworthy as possible.

An AI ethics checklist can be very helpful before developing any kind of AI system, whether it is one that decides who gets a loan, or one that answers emails. A typical checklist covers a wide variety of ethical considerations. For instance, when collecting data, is there a way to anonymise personally identifiable information? Similarly, is the data being stored in accordance with both international regulations like GDPR, and local data protection laws? A robust ethics checklist is the first step towards building ethical software.

4. *Frameworks for Navigating Ethical Dilemmas*

In the final talk for the day, Ms. Srilakshmi Subramanian, a Distinguished Member of the Technical Staff from Wipro India enumerated some of the many ethical frameworks that can prove helpful while navigating ethical dilemmas in software development. Different ethical frameworks may be relevant to different scenarios. As well, more than one ethical framework may be helpful in finding solutions in a given scenario.

Utilitarianism, or consequentialism ensures that the ethical choice is the one that leads to the best overall outcomes in terms of happiness, well-being, and/or utility. Simply, the ethical decision is the one that results in the greatest good for the greatest number of people. Software developers may use the utilitarian framework to make decisions around features or changes that maximise user benefit. An e-commerce application, for instance, might use a recommendation algorithm that promotes the most popular products, which could maximise the benefits for both the consumers who buy and use these products, as well as the business, which ultimately ends up making some profit from the sales.

Deontological ethics is another ethical framework which focuses on following moral rules or duties—the ethics are 'duty-based'. The focus is on respecting individual rights, following ethical codes, and upholding moral duties, and actions are deemed 'ethical' if they align with the principles, immaterial of the outcome. In software development, this comes into play when having to adhere to certain ethical standards, or legal regulations to protect user data, even if this ultimately results in less business profits.

Virtue ethics, on the other hand, asks what a morally virtuous person would do in a given situation, while also emphasising typically virtuous traits like responsibility, integrity, honesty, and empathy. Virtue ethics in software engineering gives developers an opportunity to focus on making transparent and fair decisions in their work, especially with respect to decisions surrounding data protection, fairness, and bias. Design-centric software that emphasises user-centric features, including better accessibility is a good example of virtue ethics.

Ethics of care focuses on people's relationships and connections with each other. With care and compassion at its core, it values empathy and the general well-being of others. In software development, this could translate into better understanding how certain decisions may affect others, particularly in vulnerable communities. Examples include apps for people with disabilities, and meditation and mindfulness apps aimed at people struggling with mental health issues, both of which focus on inclusivity and respect.

The Rawlsian theory of justice, which emphasises fairness, equity, and justice is another ethical framework that software developers can use, especially in the context of ensuring that the products they are building treat all users, especially the vulnerable and disadvantaged, equitably, without perpetuating inequalities. Unbiased and transparent recruitment tools that ensure fairness in hiring practices are an excellent example of how this framework can be applied in real life software applications.

The social contract theory argues that either implicitly or explicitly, people's moral and societal obligations are a contract between the people in order to form a society. It revolves around the idea even at the expense of sacrificing certain rights, societal rules are created for mutual benefit and order. In software development, this means that developers have a responsibility to uphold society's existing norms and regulations by adhering to local (and sometimes international) laws like GDPR to protect people's inherent rights.

Under the stakeholder theory, all ethical decisions should account for the interests of all stakeholders involved, ranging from the users to the stakeholders to the society at large. When developing software, developers should consider the rights of all stakeholders to ensure that no one group is unduly harmed. The product or software does not exist to solely increase profit, but rather to add value to the users, the creators, and society as a whole.

The precautionary principle argues that it is always better to take preventive action in the face of uncertainty, so as to curb potential harms, rather than to wait for evidence that the harm will occur, lest there be unknown or irreversible consequences. For example, before developers design new facial recognition software, they may be asked to consider the potential privacy and bias implications of such software, especially on certain minoritised communities.

Human-centred design is an ethical approach that emphasises user empowerment and experience by focusing on the needs, preferences, and well-being of users. Under this framework, users are actively involved in the design and development of the software, constantly providing feedback, to ensure that it meets the users' needs. Subsequently, human-centred design also involves testing the software with diverse users. Special fonts that help dyslexic users read better, as well as software features like screen readers, and text-for-image features are all excellent examples of human-centred design.

Towards the end of this session, the participants also had the opportunity to participate in a robust discussion about using different ethical frameworks to evaluate real world software problems. This not only gave them the opportunity to understand how to use different frameworks in different scenarios, but also helped them make actual ethical decisions in software development by combining some of these frameworks together.

Conclusions

Over the course of a day, through a series of interactive sessions with experts around the world, workshop participants learnt to better understand ethics generally, as well as the very important role it is going to continue to play in software development, especially as we move towards greater automation.

The integration of ethics in software development must be treated as a foundational aspect, not an afterthought. Ethical considerations should be incorporated into the design phase, with mechanisms in place for testing and corrective actions. Key challenges include addressing bias in data, understanding cultural dimensions, and developing adaptable frameworks for ethical decision-making. The idea of 'Ethics by Design' involves separating ethical rules from functional code to allow for flexible updates without disrupting software performance. Guardrail architecture, where ethical constraints prevent misbehaviour, is a promising approach to ensuring software aligns with ethical standards.