NAC Executive Insights

Technology
September 19, 2019

Artificial Intelligence Ethics in the Project Management and Civil Engineering Domains

Key Points

- Artificial Intelligence (AI) ethics is considered for both management of large complex projects and civil engineering.
- Broad categories of ethical concerns are explored.
- Key questions (over two dozen) related to AI ethics are laid out.

Introduction

Artificial Intelligence (AI) enabled systems, machines, and algorithms undertaking cognitive tasks raise a myriad of ethical issues. The primary concern is to ensure that AI does not harm, directly or indirectly, humans or the environment.

The perspectives in this NAC Executive Insight are twofold:

1. Management of large complex projects and the issues associated with use of the predictive capability of AI, primarily machine learning.
2. A civil engineering perspective, where AI may be employed in design and other optimizations.

A recurring question about using AI by project managers and engineers: Should we require AI ethics just as we require engineering ethics for engineers?[1]

The broad categories considered in this Executive Insight include:

- Completeness of AI ethical considerations
- Quality and limits of training data
- Hidden biases
- Confirmation of appropriateness of use for selected AI
- Diagnosis vs. design
- Accountability for AI impacts

[1] This question and other related ones are being debated today around projects, taking place under the auspices of the IEEE Standards Association and their Global Initiative on Ethics of Autonomous and Intelligent Systems that aim to address ethical issues relating to the creation of autonomous and intelligent systems. Their good work is not repeated here.

- Validation and verification
- User data rights

Completeness of AI Ethical Considerations

While the broader field of AI is placing greater attention on AI ethics, these considerations are receiving inadequate attention in the areas of project management as well as in the civil engineering design space.

The challenges posed by ethical considerations arise in project management as predictive analytics moves beyond prediction towards optimization of execution and recovery plans. Do the optimization algorithms AI enables take sufficient account of various areas of social responsibility? Is optimization merely around first cost and schedule or is it around life cycle performance in cost, environmental, and social dimensions?

Similarly, does AI enabled design optimization sufficiently consider safety during construction as well as operation and even eventual facility decommissioning? Does it consider a broad range of operating scenarios or environments or is its intended use-case narrower than what we may perceive?

Operating systems such as water and wastewater treatment systems optimized by AI must understand the potential wide range of implications for public health and safety as well as environmental impact from operating environments outside both the training data and optimization scenario selected.

We have begun to think about some of these ethical considerations with respect to autonomous vehicles, but they grow in importance as our roads become more intelligent in their own right as well as active, system-level participants in autonomous transportation.

Each use of AI requires a thoroughness of understanding of potential ethical issues that may arise as well as an agreed to quality level of the due diligence we undertake in this regard.

One last point is worth noting. To the extent to which human actors select the AI enablement to be used, they must do so carefully while understanding its appropriateness and limitations of use. Should AI enabled programs and systems be required to confirm their level of “fit for purpose”? Is the particular application well bounded by the training data and selected optimization? Is it a reasonable extrapolation? Or is it a case of trying to use a tool for other than its intended purpose like using a hammer to set a screw.

There needs to be a formal impact assessment of potential ethical issues arising.

Quality and Limits of Training Data

The quality and limits of the training data used to initially develop and tune the AI algorithms require significant consideration in order to minimize the potential of some of the ethical issues described in the prior section.

The subject of bias is covered in the next section, but is part of the data quality assessment.

Training data must be of high quality with definitional consistency and within a well understood context. Data from within singular enterprises with vigorous standardization of overall execution approach (well defined processes and procedures) may provide the highest levels of predictability confidence, but may be more limited in their ability to predict performance in other similar enterprises or even different business lines in the same enterprise where execution methodology may substantively differ (energy & chemicals vs. infrastructure).

The inherent limits of the training data also need to be understood. In the case of project predictive analytics, this may be size range represented by the training data or complexity, to name just two. In the case of design, upper limits on extreme events may not support extrapolation beyond the training data's range, especially for nonlinear performance.

Data integrity represents another important consideration. Do training data accurately portray actual project performance or are the critical initial start-up and ramp-up months reflecting plan data in the absence of effective project measurement and data capture? Do we need clear standards on data to be used in AI and would domain-specific data ontologies be beneficial?

With respect to design focused algorithms, do training data have a bias towards one particular measurement parameter versus lower quality data concomitantly collected?

Data integrity also requires understanding to what extent the relevant data environments or measuring protocols differ across the training data.

Data must go beyond addressing the concerns just outlined, ensuring that the algorithm has access to sufficient meaningful data to derive appropriate algorithmic conclusions, but also that the checking data are drawn from the same sample. In one predictive project analytics effort, the training data encompassed 70 projects with the checking data drawn from the same pool representing another 30 projects.

In designs focused on predicting behavior in extreme events, special challenges exist in using sets of extremes from multiple sample sets and recognizing that data fit on the right tails may not be as neat as modeling may suggest with the tails being significantly fatter.

In both the case of projects as well as in the case of design-based algorithms it will be increasingly important to include relevant "dark" data. An example of dark data can be illustrated in the case of project predictive analytics where including only direct project data may predict symptoms but miss the driving "disease" causes that may come from events external to the project itself. Our tendency to consider projects and other design problems to be well bounded may act to introduce an optimism bias in the results.

The ethical use of AI requires us to not suspend judgement. The results must "feel" real and believable. Validation and verification are discussed in a later section of this Insight. Results which seem counterintuitive or unduly minimized or inflated must cause us to look first at the quality of the data we have trained on. The veracity and quality of the results flows from our initial data sets.

When AI algorithms produce errors, and from time to time they will, it is important that our diagnosis of errors include a review of the quality and limits of the training data.

Hidden Biases

Bias, especially hidden bias, in our training data and the derived algorithms represents a special ethical challenge for both project and design deployments of AI. Much has been written about inherent bias embedded in various human resource and credit systems, where past human biases reflected in the training data become embedded and even reinforced in the developed algorithms. In effect bias is perpetuated in a system where no social bias is desired.

We must minimize or better yet, eliminate, human, algorithmic, or embedded data bias. AI can learn and reinforce any bias present, and these efforts must begin with understanding and monitoring training data for hidden bias.

In the deployment of AI to project predictions, different types of bias become important to discover. These include:

- Data availability bias – selecting project training data only from well documented projects even when the biggest failures were not as well documented.
- Data myopia – selecting only readily available data even when closer inspection may suggest "dark" data is a principle influencer.
- Stereotyping – classifying a member of the data set or a project to have similar characteristics to other projects without adequately confirming. Similarity of projects and project execution systems become a key consideration, creating either desirable or undesirable bias depending on intended use.
- Confirming bias – selecting only data which seems to fit our preconceptions
- "Not invented here" bias – resistance to use tools, data, or knowledge developed outside the specific enterprise

Similar biases are applicable in a design environment together with:

- Congruence bias – which limits consideration of alternative hypotheses to the one we have set out to test.
- Anchoring bias – first thoughts or information shapes decisions and thinking.
- Status quo bias – tendency to maintain current approach even when better choices are apparent ("we have always done it this way").

Are hidden biases misleading us, allowing us to feel comfortable with the very outcome we are trying to improve upon through the use of AI? Are these hidden biases embedded in our data and algorithms sufficiently evident so the limitations of our AI deployment are readily understood? Do biases reflect only one desired optimization point (cost or time) while sub-optimizing other key points such as health, safety, environment, or sustainability?

We require insight into AI optimization parameters.

Confirmation of Appropriateness of Use for Selected AI

We have already touched upon the ethical dilemmas created by using AI for other than its intended purpose. In a predictive project setting it may lead to taking corrective actions, often inadequately planned and analyzed, where none or completely different ones are required. Beyond the direct and indirect impacts on project performance, it may lead us to a less than complete overall optimization as the AI focuses on cost or schedule. This may be to the peril of broader societal consideration around health, safety, and the environment.

In a design setting, the consequences risk being even more severe, perhaps leading us to believe in the safety of a structure or process for which the algorithm has never been adequately trained and tested. As we bio-engineer new agents for water or waste treatment, we may find their behaviors and properties to be outside the testing data parameters and range.

Confirmation of appropriateness of use should consider:

- Assumption tracking and linkage to AI use cases
- Constraint awareness and tracking as it relates to the AI we deploy

Is a desirable feature of AI enabled programs to test the fit and appropriateness for the use-case at hand? Should use-case “scoring” be a feature we require in the AI we employ in both project and design environments?

Diagnosis vs Design

Understanding how developed AI will be used is key in determining fit for purpose and appropriateness. For example, diagnosis must have high confidence to reduce the level of false positives. This is similar to what we try to achieve in predictive project management analytics. Signals that predictive analytics detect can have less to do with the negative outcome than other factors such as characteristics of the company, selected approach, or execution sequence.

The consequence of misdiagnosis must be considered. As an example, Google’s Deep Mind produced a high confidence level of 94.5% correct on 50 common eye problems. While a 5.5 % false positive level may be acceptable in diagnosis on common medical maladies or even in project failure predictions, it is not in design.

Design cannot accept failure with impactful consequences. What is an acceptable confidence level for safety and how does it compare with designs developed today by licensed engineers?

Accountability for AI Impacts

Accountability is required as part of any ethical system. This must include individual accountability when considering AI impacts on society. This accountability cannot stop with just the companies. This ethical accountability by the individual is essential, given the inherent opaqueness that will persist in AI enabled algorithms despite our strongest commitments to transparency.

Where does this accountability end? Clearly it must end when AI is deployed for use for other than its intended purpose. This further elevates the need for transparency of training data and importantly, clearly defining its range of applicability and limits on the use-cases bounded by the training data and any subsequent data additions from which the algorithm continues to learn.

The need to document the design and assumptions is high and the requirements for transparency and traceability even higher.

In design deployment of AI, the safety of a system must consider the potential for multiple instances of AI algorithms running in tandem or parallel, rather than verifying system behavior in all operating contexts. System behaviors well outside our design parameters may now be possible. The same is true in utilizing different (or even the same) AI optimizations of design, where the boundary/interface conditions may be well outside those assumed in any single optimization.

A key question related to ethical accountability is whether AI is making decisions on its own or in conjunction with humans. If it is acting independently, which norms are guiding it? Will it sacrifice one individual to prevent a broader disaster or save one even if a broader disaster may result? When AI fails, responsibility must be clear. Who is responsible for a mistake?

Value alignment becomes central and human-centric AI must align with the values and ethical principles of society while demonstrating sensitivity to a wide range of cultural norms and values. Should AI be shaped to bear public safety in mind (safety, environmental, and social impacts; resilience)? When AI takes on a cognitive task previously performed by humans, does it inherit the social requirements?

What new obligations are created in the use of AI in design or project prediction? What new or improved benefits to society does AI enable?

One final point on accountability: transparency and caution are required when AI is used on cognitive tasks with social dimensions. Predictability is important.

Validation and Verification

Validation is the process of checking whether a specification, or in this instance a use case, accomplishes the stated purpose it is intended for. Verification is the process of confirming that the AI supported algorithm meets the specification and supports the use-case. The use-cases can include high confidence predicting of projects likely to fail earlier than the project manager is able to otherwise recognize this potential extreme outcome or optimization of a complicated element of design, such as minimizing weight in an aircraft bulkhead.

As AI permeates all aspects of the management of complex projects as well as their design, it is important that we have high confidence in their behaviors to avoid individual, enterprise, and societal harm.

Verification and validation are independent procedures that are used together for checking that a product, service, or system meets requirements and specifications and that it fulfills its intended purpose. These are critical components of a quality management system

Validation must assure that the AI embodiment meets the needs of the customer as well as other identified stakeholders, such as we see in the broader ethical responsibilities we assign to engineers and other licensed professionals. Independent validation and subsequent certification will become increasingly important. Project and design professions must define standards for independent validation and certification, including descriptions of limitations on applicable and appropriate use-cases. Validation must answer the question of whether we are "building" the right thing. What constitutes intensive validation of algorithms is not well defined. Do we need a neutral, secure testing environment with validation certification?

Verification, or the evaluation of whether or not a product, service, or system complies with a regulation, requirement, specification, or imposed condition is often an internal process. Verification standards must exist to ensure that deployment of AI is within the bounds of its validation and intended use. To the extent that AI enablement's can self-verify or score their fitness for purpose when applied to a specific use-case, concerns about unethical use will be somewhat mitigated. Verification answers the question of whether we have built it (or are using it) right.

Verification may be difficult, if not impossible, with some AI instances, but this is where we must at least strive for explainability and reasonableness of outcomes or predictions.[2]

The robust validation regime that is suggested in order to achieve a degree of confidence in the validity of the AI algorithm and therefore confidence that it has been represented and used with full cognizance of ethical considerations necessitates a high degree of explainability. Explainability in AI should have as a minimum the following attributes:

- Decision making process should be explainable.
- Recommendations should include sufficient explanations, data used and limitations, reasoning.
- AI decision processes should be verifiable.
- AI intent should be transparent.
- AI algorithms may be powerful and scalable, but also transparent to inspection.
- Understand how added data change expected outcomes.

Finally, an AI validation process should consider the safety of AI algorithms and be confident they are predictable in a given instance, even if the AI behavior is not.

User Data Rights

User data rights is an area of emerging concern. Individual users who contribute data to a multi-enterprise training data set must retain sufficient rights over their data while the broader (multi-enterprise) insights gained are derivative. User data must be protected and users must maintain control over access and usage of their data.

Other data, of uncertain or unknown provenance, should not be used in an AI algorithm or service without confirming rights, applicability, and appropriateness for intended use and embedded bias in data.

Users should ensure that both their data and any AI algorithms they make available are robust against manipulation.

[2] One useful technique is to look at extremes of various variables (0; ∞) and assess directional reasonableness of outcomes. This is akin to squeezing a toothpaste tube, something is going to come out.

Potential AI Challenges with Ethical Implications
• Algorithmic bias (Lack of transparency about what goes into AI algorithms).
• Systems incorporating AI migrate in ways we don't fully understand or no longer represent original intent.
• Software designed and tested in one environment (company) on one data set risks faltering when used in other circumstances (Need to use data from several companies).
• Required cross discipline development and testing not adequate.
• Training data not sufficiently diverse (overweighed towards one outcome).
• Fragility of AI systems not appreciated. - Good for intended use supported by training data. - Assumed to be smarter than it is.

Conclusion

AI offers great potential in both the project management and civil engineering domains. With that potential comes the need to ensure what we do, how we do it and, importantly, how we represent it to reflect the core ethical beliefs of society and the respective professions. The ethical challenges we face begin with the data we select to train the AI algorithms we develop. It continues, with transparency, through our validation that the AI algorithms are fit for purpose and verification that the use-cases we apply them to are appropriate and well bounded by our training data, which should be bias free.

About the Author

Bob Prieto was elected to the National Academy of Construction in 2011. He is a senior executive who is effective in shaping and executing business strategy and a recognized leader within the infrastructure, engineering, and construction industries.