



# Human Factors in NDE 4.0 Development Decisions

Ripi Singh<sup>1</sup> · Vaibhav Garg<sup>2</sup> · GPT-3<sup>3</sup>

Received: 16 July 2021 / Accepted: 11 August 2021 / Published online: 18 August 2021  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Stage-gate process is a project management technique that has been used to manage major technology development and deployment projects for over 50 years. It emphasizes quality, risk, and value, and provides an orderly framework for decision making with multiple projects carried out concurrently. This paper will examine the potential application of this process to NDE 4.0 developments. This paper will also explore the factors that make a stage-gate process successful, and how biases affect the outcomes of gate reviews. It explores which biases are problematic and how we can de-bias our own interpretation of them through conscious awareness and re-thinking the Stage-Gate Process. It also discusses how to create an environment where biases are minimized or eliminated. During research and reporting of this manuscript, the human authors synthesized their decades of experience in engaging with technology development decisions, with contributions from an artificial intelligence agent GPT-3 on human biases and motivations. This was done to capture the firsthand experience of collaborating with AI in the spirit of increasing the role of machine intelligence in human-machine systems, typical of NDE. Since this appeared to be an innovative approach, the authors decided to use the gated process in creating this paper with AI as a co-author. The details are presented as a case study within the paper. The topic is important because, managers are struggling to embrace the NDE 4.0 initiative, despite its importance and their willingness. Leveraging AI partnership to even address this important topic is a testament to cognitive capability of learning machines to help inspection systems with creating test procedures, making decisions, and creating test reports. This manuscript is the *first attempt* at qualitative treatment of human factors in gated innovation process with research assisted by an Artificial Intelligence based language model (non-human). The first paragraph of the abstract above is completely AI generated.

**Keywords** NDE 4.0 · Innovation · Industry 4.0 · Stage-gate · Phase-gate · Gated-reviews · Artificial Intelligence

## 1 Introduction

The digital transformation of NDE industry is now quite inevitable [1]. This would involve significant innovations and forays into uncharted territories, with technical, technological, ecosystem-level, financial, ethical, and societal implications. Impactful Innovation always comes with a degree of uncertainty and risk depending upon the ‘scope of the leap’ and its impact in creating a novel solution. Revolution is an even bigger leap and brings along a much stronger

ambiguity and complexity. We can see some major winners and losers in the marketplace in first two decades of the twenty-first century. Many more will run out of steam with legacy ways and offerings, mindset of the 3rd revolution, unable to overcome fear of failure associated with the 4th revolution. Risk mitigation in NDE 4.0 technology development and adoption thus becomes an all-important task for any business in the inspection eco-system. In fact, innovation is becoming a serious discipline as evident from for emerging series of ISO standards [2] on innovation management guidance, which can be used for NDE 4.0 [3].

Amongst many tools, stage-gate reviews are a very popular processes to identify and mitigate the risks in a progressive manner for any new initiative, product, service, business model, or even strategy. Having worked with major manufacturers, the human authors have had the good fortune of being a part of several reviews at various levels and disciplines—design, technology, component, system, product,

---

✉ Ripi Singh  
Ripi@InspiringNext.com

<sup>1</sup> Chief Innovation and Strategy, Inspiring Next, Cromwell, CT 06416, USA

<sup>2</sup> Genus Power Infrastructures Limited, Jaipur, India

<sup>3</sup> Generative Pretrained Transformer Language Model Generation 3, Open AI, San Francisco, CA, USA

manufacturing, process, program, business, and so on. There are well-documented practices which make the process successful and can be adopted by everyone. Yet the maturity of the process varies from organization to organization, mainly depending upon the cultural attributes and human factors. The authors are convinced that the value of review gates comes from the review team, more than the review process. At times, the team can make a good decision despite insufficient input and significant uncertainty; whereas other times, it makes a poor choice despite clear evidence that might be contradictory.

It cannot be over-stated that a robust evaluation of programs and projects is essential for success of any NDE 4.0 initiative, and the consequences of not getting this right are dire. A serious *implication* is continuing with less promising projects with increasing demand on resources, which leads to consumption of innovation portfolio budget/resources for other projects knocking at the door, further upstream in the pipeline.

In this paper, the authors recap and customize the stage-gate process for NDE 4.0 and discuss what makes it successful and then research on what makes it interesting. They have documented several human factors and organizational dynamics issues that cast a shadow on the data in front of the decision makers, which become clear, sometimes too late. The success hinges on the review team, their willingness to surface conflicts, ability to set aside the emotions, amicably resolve difference, and make the difficult choice under uncertainty or assumptions yet to be validated.

### 1.1 Collaborating with Artificial Intelligence

The human authors initially chose to collaborate with an AI agent as a co-author to study human factors in stage gate process, in the exploratory spirit of “Why not?” [4] but quickly saw some very interesting parallels with NDE. First, every NDE method is a quality gate process with humans interpreting audio-visual stimuli/data from a machine and making a go/no-go decision. The human authors surmised that the project gates will likely suffer from similar human factors as an inspector making a decision. Secondly, the NDE 2.0 and 3.0 has been a man-machine collaborative activity with role of machines increasing continuously [5]. NDE 4.0 is now bringing cognitive capability of the machine ‘in the loop’ with a potential to completely take the human ‘out of the loop’. AI is on the cusp of becoming an integral part of inspection procedure development, data interpretation, and report creation. Finally, the human authors wanted to experience this collaboration first-hand to investigate and report on a sensitive topic and understand if the stage gate process can be used effectively for adoption of AI/ML into NDE process.

In the process of collaborating with an open mind, authors also discovered that GPT-3 helped reduce their biases by

assisting in research, writing prose, and bringing out human factors that otherwise were likely to be missed or omitted. It convinced the human authors of efficiency, effectiveness, and accuracy of collaborative effort.

The AI model was treated just like a co-author, with inputs assimilated seamlessly into the written text, along with additions and updates, just as the human authors added onto each other’s contributions in multiple iterations. The outline was developed jointly by all the three authors, with the very first machine output found adequate but not fully satisfactory. The section on stage gate process design and best practices is heavily backed by personal experience of human authors. The section on cognitive biases is almost completely written by the machine author GPT-3, with human edits constituting less than 10% of text. All the final revisions to the manuscript based on reviewer comments were addressed by the human authors.

The details are presented as a case study at the end of the manuscript.

## 2 Stage-Gate Process Design for NDE 4.0

A quick story to illustrate the point. After a nasty accident at an intersection, city authorities asked, “*Why do we not have a STOP sign here?*” Reply came back as “*People were complaining about seconds lost due to ‘stop and go’ on an intersection with virtually no traffic. Since we have not had an accident here in a long time, we took the sign off, just yesterday.*” There is a reason for a stop sign at an intersection—stop and look for what is coming from the side.

Gated process is just that. A momentary pause to become situationally aware of what is moving and may be on a collision course with your project plan. It goes by various names such as Phase-Gate, Stage-Gate, Stage-Kit, Passports, and Toll Gate process.

### 2.1 Objective(s)

The primary purpose of a gated process is to align all stakeholders and reduce the risk of continued investments. It aims to improve innovation effectiveness by separating project leadership from resource decision-making to avoid conflicts of interest; formalize points at which discontinuation decisions can be made; and nudge leadership to critically compare projects with others.

For NDE Business Objectives and Leaders: Annual or quarterly strategic reviews serve this purpose, to pause and look for where the competing technology might be coming from, and what the other PESTEL (Political, Economic, Social, Technological, Legal and Environmental) trends are looking like. All it takes is a little time with the

leadership to properly re-position for success and mitigate business risk.

For NDE 4.0 Projects and Innovation Managers: Milestone based progress reviews serve this purpose, to pause and look if it still makes sense to continue into the next major phase. All it takes is a little time with experts to properly assess technology readiness to move on and mitigate quality and development risk.

## 2.2 Model for NDE 4.0 Initiatives

The process of converting an ‘idea into value’ or ‘developing solutions to meet a need’ can be complex and chaotic. Stage-Gate model simplifies it by defining a series of stages separated by gates. Stages are where project team conducts activities and Gates are where review team makes a Go/NoGo decision. For NDE 4.0, authors propose the following as a baseline, which can be customized to meet the needs of individual organizations.

- A. Pre-development activities,
- B. Development activities, and
- C. Deployment activities

The model is designed to portray all project activities as a visual pipeline where the Gates serve to surface and prioritize the opportunities deserving of the organization’s scarce resources, acting like a filter. Refer Fig. 1 for details.

### 2.2.1 Predevelopment Activities

This includes all activities required to build a business case and help leadership decide if a particular NDE 4.0 project deserves serious attention or not. The gates [6] may include:

- (a) Purpose Clarity or Strategic Foresight: Do we understand the digital use cases of NDE 4.0 [7] and agree

- with their alignment with the overall technology road-map and growth strategy?
- (b) Market Insight: Do we understand the changing customer (user or inspector) needs or market gaps, competitive forces, and regulatory or compliance requirements?
- (c) Technology Options and Ideation: Do we have enough ideas and digital technology options that can be integrated into a cyber-physical loop [7] to address the market insight?
- (d) Value Proposition: Will the cyber-physical loop create value in the NDE 4.0 eco-system [8]?
- (e) Concept Qualification: Can a promise of digital transformation be delivered profitably with current or even a new business model?
- (f) Resource Approval: Project has a strong business case and has been prioritized to initiate.

### 2.2.2 Development Activities

Generically speaking, all qualified concepts become eligible for resource allocation and should feed the pipeline for development activities. In a strong organization, there should be enough qualified concepts that leadership can choose from and build a portfolio of projects to maximize returns and manage the risk. The first gate would then be to get funded. In a sense that is not a gate for the project but a gate for the business to make strategic investment decision. Once the project is funded, it may go through technical, marketing, and operational gates.

The authors believe that Technology Readiness Levels (TRL) used by DoD/NASA [9, 10] and many other commercial organizations is among the best practices. A similar one called Manufacturing Readiness Levels (MRL) is used by many organizations [11]. Although these models were developed during the third industrial revolution, they are valid in the digitalized world as well from process perspective. They do however require tweaking and interpretation

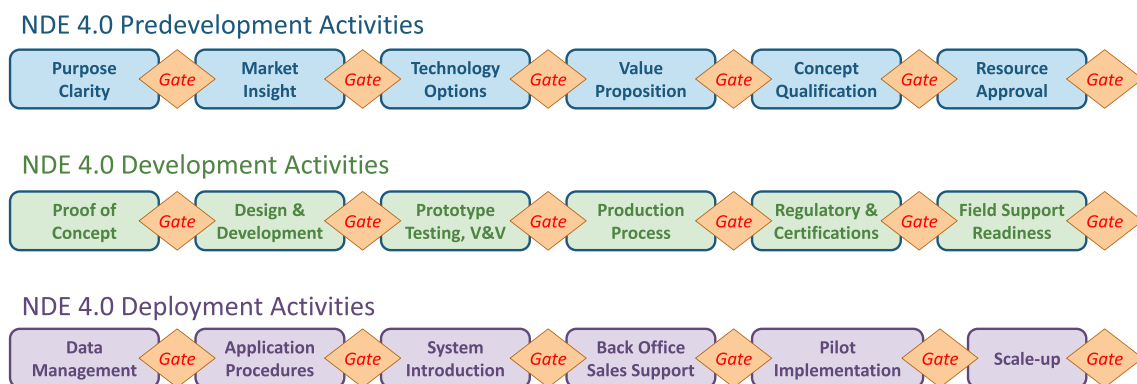


Fig. 1 Graphic representation of a typical stage-gate process for NDE 4.0 initiative

that is more attuned with the digital transformation, connected world, and an eco-system perspective. If you find TRL/MRL process to be too complex to deploy or adopt, a simpler sense could be defined as:

- (a) Proof of Concept: Do the core concepts work in lab environment and are they in line with NDE 4.0 principles [12]?
- (b) Design and Development: Do we have enough details to produce the physical product, (or service), digital system including IIoT, Digital twin, Data structure, Data flow and integration, visualization?
- (c) Prototype Testing: Does the full up cyber-physical loop work in real environment as expected?
- (d) Production and DevOPs: Do we have the setup to mass produce the physical system, integrate with software, and deploy the software in the field ensuring scalability and Service Level Agreements for reliability and availability?
- (e) Regulatory and Certifications: Have we received necessary approvals to sell and deploy the system alongside with human operators?
- (f) Field Support Readiness: Have we developed the necessary items required to support the users of the system—Training, System (re)calibration, semantic interoperability, data security and integrity, application guidance etc.

Depending upon complexity, confidence, and risk assessment, the project may have more or fewer gates. A simple study project may have an interim content review and final report and review. Complex product design may have a hierarchical gate structure such as component design gate(s) to support a system-level design gate.

### 2.2.3 Deployment Activities

Once the system has been validated and certified, it is ready for deployment in field. The gates may include:

- (a) Data Management Options: Are there novel ways of engaging with the customer, through use of digital twins/threads and data based value creation?
- (b) Application Test Procedure: Have we figured out the test procedures for the customer application and customer education/training needs?
- (c) Systems Introduction: Do we know how to introduce the system into existing infrastructure?
- (d) Back-End Support Readiness: Do we understand the technical nuances of data management and analytics. Is it compatible with existing operating model or do we need a new model?

- (e) Pilot implementation: Have we acquired and learned from the first few applications, particularly the data use for systems assisted decisions?
- (f) Scale-up: Expansion to full scale deployment, training, and assimilation.

Deployment activities do not need to wait for all development activities to finish. It generally starts before the lead users are engaged. In fact, in some cases the application procedures might be required for certification during development activities. Also, development activities may include steps to manage calibration, maintenance of equipment, and even end of life, such as recycling/repurpose etc.

### 2.3 Outcome(s)

The outcome of each review gate could be categorized under various classes:

- GO: The team is authorized to proceed to the next stage. There may be reviewer comments or recommendations for the next stage.
- Conditional GO: The team is authorized to proceed to the next stage, with very specific actions items that must be completed by a specific date (ASAP) to the satisfaction of the review team, for it to be considered a GO.
- NO-GO—RE-DO: The team is not authorized to continue to the next stage at this point. They are assigned specific action items and asked to return for another review. This is an indication of serious gap between project parameters and the gate success criterion.
- NO-GO—HOLD: The team is not authorized to continue to the next stage at this point. There is a specific unmet criterion, or changed context, which requires a temporary hold of the project. The hold may be until the criterion is met or just a for a specific period, at which point it should be brought back for review again.
- KILL: Completely abandon all effort for ever and let the team move on to other activities.

The organization can choose to have additional possible outcomes or have different names for these. The point is that there should be enough options to minimize both rework and schedule delays and avoid bigger cost implications later.

#### 2.3.1 Implications of Various Outcomes

The purpose of a review gate is to reduce the risk, by ensuring continuation if the concept still qualifies. The review process is susceptible to human factors and the review team can make an error in judgment, despite all the data and experience. For 5 possible outcomes defined above, there are 5 possible True outcomes and 20 possible False

outcomes. Each outcome has a different impact. To reduce complexity, let us just limit ourselves to Go and NO-GO options. Figure 2 below presents the implication of human error at the review gate. Notice the table is similar to NDT outcome, which in a sense is a quality gate for the artifact under evaluation.

A *true positive* (GO when it should have been a GO) ensures progress, confidence, and team buy in.

A *true negative* (RE-DO when it should have been a RE-DO) improves learning, saves failure later reducing losses, builds stronger teams, and drives humility.

A *false negative* (RE-DO when it should have been a GO) leads to some unnecessary rework, schedule delays, and a demoralized team when they are confident. The positive that comes out is improved communication.

A *false positive* (GO when it should have been a RE-DO or a NO-GO) will continue to accumulate losses and even an escape of a poor design to the market leading to liabilities. They usually make good learning stories, which drive process improvements, policies, and even regulations.

You can observe that there are both positive and negative implications to both correct and incorrect review outcomes.

### 3 Practices for an Effective Stage-Gate Process in NDE 4.0?

Thamhain [13] makes a case for careful integration of the stage-gate process with the various physical, informational, managerial, and psychological subsystems of the enterprise and its cultures and values, to be effective. The human authors agree with many of his observations. An important one being that the team members must work in an environment conducive to mutual trust, respect, candor, and risk sharing. Equally important, the work environment must foster effective communications, cross-functional linkages, and a business process conducive to interconnecting people, activities, and support functions.

The organization should facilitate certain practices to make the stage-gate process effective and efficient. Let us look at these controllable factors first.

#### 3.1 Success Criterion

Each gate must have a clearly defined set of criteria for a GO, based on the activities of the last stage. Meeting these criteria should essentially ensure successful completion of the project. The criteria must be agreed upon by all the stakeholders at the beginning of the project cycle, with specific metrics wherever possible. It should be transparent to the project team, and they must agree to it at the previous gate.

#### 3.2 Review Team

The gate reviewers should be assigned early in the project life cycle and fully committed to the success of the organization, and not biased towards the project. They must represent all disciplines appropriate for that gate. The suggested participation for the project review team under stage-gate process for NDE 4.0 initiatives includes the following roles:

- Funding sponsors, who are accountable for profit & loss.
- Product/service line heads, who will eventually own this system.
- New Business Development, which must generate revenue from this.
- The project team, to defend the progress and learn from gate experience.
- NDE method and Digital technology experts, who are responsible for technical excellence.
- Optional invitees, such as retirees, Consultants, or Customers, for wisdom.
- Representatives from user group (who wear customers' hat such as NDE level III) and
- Support functions such as legal HR.

Fig. 2 Implications of true and false calls at gate reviews; just like an inspection outcome

Review Outcome	Given a Re-DO	Given a GO
Deserves a GO	<b>(False Negative)</b> (-) Unnecessary rework. (-) Demoralizes a good team. (+) Improves communication	<b>(True Positive)</b> (+) Ensures progress. (+) Build confidence. (+) Gets buy in.
Deserves a Re-DO	<b>(True Negative)</b> (+) Improves learning. (+) Saves bigger failure later. (+) Builds stronger teams. (+) Drives humility.	<b>(False Positive)</b> (-) Increase losses. (-) Exposes to a liability. (+) Makes good learning stories. (+) Drives process changes/policies.



Certain roles may skip certain gates, depending upon relevance, to keep the process efficient.

The members of the review team should not have any conflict of interest. The outcome of the review gate should not have any impact on their professional or personal lives, in any way shape or form.

A review team must have a chair or the primary gate-keeper. This can change as the project goes through various gates depending upon the primary success criterion. For example, Detailed Design review may be chaired by a Chief Engineer, while the pilot implementation will be chaired by ASNT Level III or equivalent.

### 3.3 Stage Execution

In an NDE 4.0 project, it is advisable that management allow as much creativity during the stages in between gates as practical. Preference should be to keep the rigor to gates. This advice must be contrasted against the prevailing wisdom in manufacturing setups or for incremental engineering projects, where gates are considered a quality assurance step and thus are considered a waste to be eliminated or reduced by putting rigor into the stages. Refer Fig. 3. Professionally interesting and stimulating work appears to be one of the strongest contributors to a successful gate later. Project leaders should try to accommodate the professional interests and desires of their personnel whenever possible. It leads to increased involvement, better communications, lower conflict, and higher commitment. One of the best ways to assure interesting work is for the manager to match the personal interests with the scope and needs of the tasks carefully during the "signing on" of personnel to the task or project team. In addition, the manager should build a project image of importance and high visibility, which can elevate

the desirability of participation and contribution. Such an environment helps to motivate people toward established goals, innovatively and creatively, and cooperation with stage-gate process.

### 3.4 Gate Execution

First thing is that a gate review needs to be taken seriously. Project team must come prepared to present and discuss. A well-organized status summary, major actions, and closure from previous gate, change announcement, or problem statement often requires detailed background work and considerable effort organizing the presentation or discussion. Sometimes merely the process of preparing for the gate review throws up concerns that the project team must address to make the project successful.

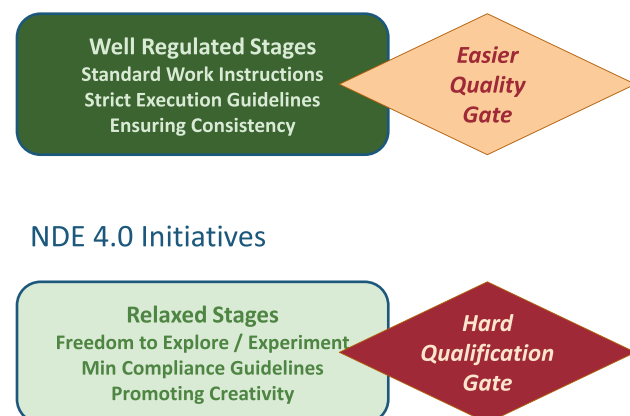
Then review team must receive the presentation and data well in advance so that they have a chance to prepare as well. The review meeting should have ample time to present, discuss, explore, and demonstrate that project meets the agreed upon success criterion to the satisfaction of every review team member. Gate decision making is not meant to be a democratic process. It is not about majority votes. Each discipline must say "Go" for it to be a go, just like you see in space vehicle launch.

At the end of the Gate review meeting, assign action items and follow up, publish major milestones, and concur with project objectives. This provides cross-functional visibility for the overall project and helps to unify the project team toward critical outcomes on schedule.

Meetings must be highly interactive and run on a mutual trust basis. Good review meetings are often noisy with plenty of candor and broad involvement. It is this dynamic which helps to discover small problems at an early stage. The review team needs to maintain a healthy level of conflict and collaboration at the same time amongst various roles/disciplines. For example:

- The development and deployment folks should collaborate throughout to work toward the same innovation and timeline.
- Talent development manager and innovation chief should collaborate for proper talent acquisition and development.
- Program manager and subject matter experts could have a perpetual conflict between product/service quality/excellence and project cost/schedule performance.
- Market insight experts and subject matter experts could either have a conflict or be collaborative depending upon customer push and pull for innovation.

#### Traditional Projects



**Fig. 3** The contrast of stage-gates between traditional projects and NDE 4.0 initiatives

Having seen so many reviews with passionate debates, leading to some exciting outcomes, the authors believe

*“conflict at the review gate is a good thing. How we choose to resolve that conflict during the review or afterwards—as an action item, defines the review gate experience and employee engagement with innovation”*. The team ought to go into the gate review with an open mind, focused on the purpose, objectives, ethics, and for educating each other. In the end, *both the project and the review teams are on the same side, addressing uncertainty*.

A project should only move forward to the next stage, when both the product/service development and business managers feel that the concept still makes sense; and all the previous gated criteria have been successfully met. Any exception to the review gate or waiver of criteria/expectations should require a review team approval. The higher the uncertainty, the lower the first pass yield.

## 4 Challenges with a Stage-Gate Process in NDE 4.0?

Like any organizational system, the gated process does not always work as intended. It is a highly emotional event, irrespective of objectivity designed in the process, using all the above-mentioned practices. The fourth revolution makes it even more challenging because most experienced reviewer teams are still thinking with the NDE 3.0 mindset. Let us look at some of the human factors that make it interesting.

### 4.1 Applicability to NDE 4.0 initiatives

The authors thank Dr. Anukram Mishra, CTO, Genus for asking an important question about the applicability of the Gate reviews to all projects. He questioned if there was a way to determine if a given project is right for stage-gate process or not? Sethi and Iqbal [14] argue that Stage-Gate controls have the potential of restricting learning in a new product development project and thus hurting the performance of novel new products. They specifically observed through data, that control on new product development exercised through rigorous gate review criteria, increases project inflexibility, which in turn leads to increased failure to learn.

With extensive role of digital counterpart in NDE 4.0, software side of development will most likely use some form of agile development contexts. In this style, the speed of execution and ease of deployment trumps rigor in gate reviews. This is probably because of the ability to pivot hard and pivot fast easily and to be able to use the early adopter customer base as proxies for the gate review committees. An example may be an exploration Horizon 3 project that seeks to establish the physics of the NDE system rather than the engineering aspects of the same. The very high uncertainty and limited visibility may warrant a “fire and forget” approach.

The right amalgamation of agile processes with the stage-gate process with special focus on NDE 4.0 might be warranted in certain projects. The authors have deferred that topic for future work.

### 4.2 Success Metrics

Sometimes, the management chooses a metric around increasing the first pass yield of gated reviews. That is not a good practice. It drives many wrong behaviors: (a) a tendency to pick low-risk ideas/projects to begin with, (b) to keep working to perfection, and (c) the review team’s bias towards a ‘GO’ outcome. That is all counter to innovation and the purpose of a gated review. You want to fail fast and learn fast. It is OK to track, but not to set a goal for yield. Another problem with this metric is that it assumes a balanced portfolio. A 95% first pass yield may mean that the top 5% of the innovative initiatives were all filtered out.

Schedule pressure creates a tendency to compromise marginal situations, and nudges towards taking decision with insufficient data or to under-estimate the risk. And if there is a sense of urgency or a need to meet a certain specific performance metric, the entire interpretation of the data gets skewed.

### 4.3 Cognitive Human Biases that Affect Gate Review Outcomes

Decision making at the gate review gets affected by bias from various directions. The reviewers can get emotionally vested with the idea and progress based on watching it evolve with their input at previous gates. Experience creates an anchor bias to previous success stories and traditional ways of doing things. Individual biases [15] also play a role in team dynamics if the review team chair is not able to handle conflict.

Biases and conflicts of interest affect gate review performance by influencing the interpretation of data and the selection of criteria for success and failure. Existing bias in the review team can affect the project gate review outcomes by biasing their interpretations of the findings. This may lead to situations where objective assessment is almost impossible, and the review just manifests predetermined outcomes. In other words, they can be looking for information that will support their existing bias or what they believe about the project, instead of seeking information that will give them new insights.

#### 4.3.1 Bricklayer’s Fallacy

While diversity of background and experience in review team helps reduce certain kind of bias, a common theme to watch for in NDE 4.0 is a bias towards the proven ways of

doing things from the third revolution. For example, using digital and physical systems development as separate entities and trying to force fit solutions to existing standards and certifications. The example of rear-view mirrors for a driverless car.

The cognitive bias to fall into the NDE3.0 thinking while working on NDE 4.0 transformation is called the “Bricklayers’ Fallacy”. It is the assumption that the context will remain the same as the future unfolds, and therefore, that it is unnecessary to build into the new paradigm any new means to support the growth into the future. Given that the field is likely to evolve rapidly, an approach based on the YAGNI (You ain’t gonna need it) approach may not be the right strategy. It would be preferable to take a platform approach making room for scalability in future. To rephrase in context with NDE 4.0 transformation, bricklayers’ fallacy is the assumption that the new physical inspection technologies and data analytics will remain constant, and rather than designing for this evolution, companies should wait for it to happen.

#### 4.3.2 Stereotyping, Recency Bias and Availability Heuristics

Stereotyping occurs when one applies a mental shortcut to a problem, by using generalizations. This can lead to overconfidence, especially when information is scarce. Typically, in high risk or technologically complex project reviews, the review team tends to have a tendency to fall back onto simpler explanations and drawing parallels to their own past experiences, especially recent ones (Recency bias—the tendency to weigh recent events more heavily in memory than earlier events of the same kind, a common concern in NDE practice). While this thick slicing may be useful, it can lead to the review team disregard the nuances of the current projects in favor of—potentially—irrelevant factors.

This problem can be exacerbated by another cognitive bias called as the Availability heuristic, which is the tendency to judge the frequency or likelihood of an event by the ease with which relevant instances come to mind. In the late 1960s and early 1970s, Amos Tversky and Daniel Kahneman explained that judgment under uncertainty is often based on a small number of mental shortcuts and heuristics rather than extensive algorithmic systematic data-based processing. For example, in judging how likely an event is, one might ask, “How easily can I recall examples?” This results in the “registered” impressions from recent projects being given more weight than the more relevant ones from further in the past.

Availability Heuristic also interferes with the likelihood calculations for risk assessments. In terms of the innovative technologies and thought processes under evaluation for NDE 4.0 implementation projects, the cognitive load is understandably high and the natural fall back for most

reviewers is to fall back on their experiences which, to a substantial extent, are based on Industry 3.0 initiatives. The lens of Industry 3.0 may cause misestimation and misplaced paranoia with regards to the project under review. While this offers insights into the cognitive processes that explain human errors in judgement without invoking motivated irrationality, we need to be cognizant of this heuristic to avoid falling into this trap. In other words, while this is a natural trap to fall into without insinuating reviewers of mal intentions, it would be wise to be wary of this.

#### 4.3.3 Delicate Engagement

This is an even more serious problem if the review team members are the same people who developed the project plan or have been somehow involved in the project execution. Continuous engagement of review team with the project, even when limited to reviews, leads to a 2-way entanglement. The project team learns how the review team thinks and develops ways to influence the outcome of the gate review. The review team gets empathically attached to the project (& team) and bias towards their earlier feedback. Since the gate committee is not a machine devoid of emotions and biases (yet), they will find it incredibly hard to stop or kill a project in line with the changing marketplace environment. It is thus important to bring back a higher authority at later stage gates into the review mix, who were involved in the original approval of the project and can objectively look at the project against the original success criterion.

NDE sector places a lot of emphasis on Level III proficiency folks, and there are not enough of them to serve in different roles. This cause the above conflict-of-interest concerns to be exacerbated.

#### 4.3.4 HiPPO Effect

The most dangerous voice in a review meeting is the HiPPO: (Highest Paid Person’s Opinion). The person with the biggest salary/title can crush diversity of thought and more, if not careful. Trying to keep the big boss happy, can sway decisions in one direction. Opinions of certain well recognized individuals carries more weight than data-based evidence. Status disparities can fuel conformity and groupthink. When you need diversity of thought, ask everyone else to share their views before turning to the HiPPO.

Once again, we need to be careful about role played by assertive and experienced NDE professionals.

#### 4.3.5 Personal Insecurity

A gated process often comes across as a threat to career progression or job security. Managers feel uncomfortable at the idea of a NoGo outcome. Management must foster a



project team environment of mutual trust and cooperation, an environment that is low on personal conflict, power struggles, surprises, unrealistic demands, and threats to personal and professional integrity. After effects should not include unnecessary inferences to performance appraisals, tight supervision, restriction of personal freedom and autonomy, and overhead requirements. This is especially true for the innovative NDE 4.0 initiatives that are, by the very nature of innovation, are high-risk high-reward initiatives, where fear of failure can cripple the project teams.

#### 4.3.6 Anchoring and Adjustment

Anchoring and adjustment are a cognitive bias that describes the common human tendency to rely too heavily on the first piece of information offered (the "anchor") when making decisions. Anchoring entails that the review team's decision making is clouded by irrelevant information. When the discussion is about the project plan, the team will bring in factors such as the number of person-months for a project task. This can be an inadvertent action due to the order in which the project team presents the project progress or can even be a nefariously planted step to fixate the review team's attention to trivial aspects rather than on some of the aspects where the project team may be uncomfortable. Experience creates an anchor bias to previous success stories and traditional ways of doing things.

#### 4.3.7 Framing Effects

Framing effects occur when the same issue can be seen in different ways, leading to different choices being made. There are two framing effects that can affect the effectiveness of project gate reviews. The first effect is that people are more likely to reject a new idea when it is framed in a negative way. For example, when they are told that the new idea will cost more than the old one. The second effect is that people are more likely to accept a new idea when it is framed in a positive way. For example, when they are told that the new idea will have a higher probability of success. To avoid these effects, project gate reviews should be framed in a neutral way, with all aspects clearly represented, and a predetermined qualification criteria with weightages so that the decision is data driven. The authors have also seen cases where the salespeople (good at framing) get away with substandard ideas being funded to support their initiatives as compared to ones that may be more valuable to the company.

There is an anecdote about a frustrated Ph.D. student that illustrates this effect. The student was frustrated by the advisor's incessant and repeated requests for major revisions on each review. One time, he decided to inject a few major spelling and grammar snafus in the opening paragraph of the paper. The advisor latched on to those and spent his whole

time identifying more such errors, completely forgoing the criticism of the content.

#### 4.3.8 Confirmation Bias

Confirmation bias is the tendency to search for or interpret information in a way that confirms one's preconceptions, leading to statistical errors. Confirmation bias occurs when people focus on information that supports their beliefs, while discounting other information. This phenomenon can be harmful to project gate reviews as project managers are likely to present their proposals in a way that is most likely to be supported by the people reviewing them. Also, the review team will hear what they want to hear, not what is being presented. This may mean that project managers, being wary of being rejected offhand, are less likely to present the more difficult problems they knew existed with their proposals. As a result, these problems may not be surfaced until it is too late to make any changes.

#### 4.3.9 Probability Neglect

Project gate reviews should be completed to ensure that the project is on the right track and to identify any potential risks. The risk identification is often not done well because of risk neglect cognitive bias and because of the uncertainty about the probability of the risk. This can cut both ways. The review team can overplay a risk with low likelihood and downplay a risk with high likelihood. The review team can also underplay a risk with high likelihood and not pay attention to a risk with low likelihood. The project gate reviews should be completed by a review team with individuals who are not on the project. This would ensure an unbiased evaluation of the probability of the risks. The availability heuristic, discussed above, also tends to distort the perception of the likelihood and severity of risk, based on recent experiences.

#### 4.3.10 Sunk-Cost Fallacy

Sunk-cost fallacy is the tendency to continue an endeavor based on the cumulative prior investment, despite new evidence suggesting that the investment is likely to be wasted. Sunk cost fallacy causes the review team to keep a project going that may not be worth the time and effort required to finish it based on the time and effort that has already been spent on it. This can lead to a project review that does not make any viable recommendations for the project to continue, even if there are more viable options. Basically, the project review is not objective to the data. The decision is heavily influenced by what has already been spent to produce the project.

The way the project review teams can avoid falling in this trap is to look at the project from a global perspective. The review team should look at the alternatives available to them and ask themselves whether the project will produce adequate results to justify the time and effort already spent in the project, in addition to the time and effort still required to bring it to market.

Another way the mental effort from the sunk-cost perspective can be reduced is by making a conscious effort to identify and document the “learnings” from the project that can directly be assimilated as value-adds in other present and future projects and treat them as positive outcomes when pulling the plug. As an example, The Hubble space telescope had a horrible focusing error in the first launch, that rendered all images horribly fuzzy. The software technology that needed to be developed to correct the aberrations enabled Breast cancer detection using Mammograms.

#### 4.3.11 Social Influence and Compliance

The need for social influence and compliance has an important role in project gate reviews. The need for social influence and compliance within the reviews can be seen as a pressure, which may inhibit the honesty and transparency within the reviews. The perception of this pressure may be different from person to person and could result in a different outcome of the project gate reviews. This can lead to the review team being risk-averse and opting for decisions that are mainstream. This can lead to a Group think bias where reviewers develop an identity as a group, focused on a mission, and reinforced by rewards. There is pressure to conform, as well as a fear of dissent.

#### 4.3.12 Other Human Biases

There are a few more cognitive biases that can affect the gate reviews adversely if the participants are not careful at recognizing them and keeping them at bay.

- (1) Hindsight bias is the inclination, after learning an outcome is true, to see it as having been predictable.
- (2) Optimism bias is the tendency to be over-confident and believe that good things are more likely to happen to them than to others.
- (3) Self-serving bias is the tendency to claim more responsibility for successes than failures.
- (4) Egocentric bias is the tendency to over-emphasize our own attributes and to under-emphasize the role of external factors.
- (5) Outcome bias is the tendency to judge a decision by its eventual outcome instead of based on the quality of the decision at the time it was made.

- (6) Choice-supportive bias is the tendency to remember one's choices as better than they were.
- (7) Good-job bias is the tendency to evaluate a task by the ease with which it is performed, rather than by the objective quality of the results.
- (8) Illusion of control is the tendency to believe we can control or at least influence outcomes that we clearly cannot.
- (9) Illusion of external agency is the tendency to ascribe one's actions to other forces.
- (10) Overconfidence effect is the tendency to overestimate one's own abilities in a given field, relative to others.
- (11) Violation of the expectation effect is the tendency to expect a given outcome based on previous experience, despite new evidence suggesting that the outcome is no longer likely.
- (12) Backfire effect is when, given evidence against their beliefs, people can reject the new evidence and believe even more strongly, thus maintaining or strengthening their initial belief.
- (13) Attraction effect is the tendency to overvalue what one already has.
- (14) Less-is-better effect is the tendency to prefer a smaller set to a larger set with a higher probability of success when the costs of the sets are the same.
- (15) Focusing effect is the tendency to place too much importance on one aspect of an event or a decision.
- (16) Contrast effect is the tendency to see a comparison as more extreme when it is made against an extreme reference point.
- (17) Representativeness heuristic is the tendency to judge the likelihood of an event by how well its characteristics match those of a typical event.
- (18) Unwarranted certainty is the tendency to decide without the necessary information.
- (19) Duration neglect is the tendency to neglect the duration of an episode in comparison to its other properties.
- (20) Memory distortion occurs when reconstructions of reality get filtered through people's minds and are not perfect snapshots of actual events.

Many of you can relate to these as human factors that come into play with inspection reliability (POD studies).

#### 4.4 De-biasing our Interpretation of Biases

If biases are so prevalent in every decision making, the human authors must also be biased in their ideas and expressions in this paper. Some biases work in favor and others working against the desire to pursue NDE 4.0. To bring some diversity of thought and hopefully lesser bias, the intelligent digital co-author provided this insight on

Availability heuristic affecting evaluation and gate review of new technologies.

1. Availability heuristic causes the reviewers to fall back on their experiences which, to a substantial extent, are based on Industry 3.0 initiatives.
2. When evaluating proposals, reviewers use the mental shortcut of the availability heuristic to look for "traditional" answers. This means that proposals are more likely to be approved if they fit the dominant thinking of existing industry, even when there is merit in the new proposal.
3. The same mental shortcut of the availability heuristic is used when evaluating gate proposals. Gate reviewers are likely to be more conservative when assessing new proposals, and less so in evaluating "safe" projects that build upon existing industry practices.
4. The availability heuristic can work in the favor of new projects, but only if they are presented as "traditional" approaches.
5. The availability heuristic may result in the rejection of potentially promising innovations.
6. A gate review board that is made up of different experts and includes a representative of the new industry can minimize some shortcomings of the availability heuristic.

With regards to the biases inherent in the AI trained on human data, the authors were careful to steer clear of topics that can have an axis of bias. To mitigate, careful attention was paid to inputs, and a semantic content filter was used for outputs. As an example, all prompts to GPT-3 used the "singular they" construction to eliminate gender bias. Special attention was paid to the outputs. Given the diversity of location, age, and backgrounds of the authors, and adding the AI to the mix, the authors hope that the intersection set of the biases is small enough.

#### 4.5 Word of Caution

Authors have been managing new product development for decades now and have seen numerous variations of the stage-gate process. Simple ones may have as few as two to three gates. Complex systems have multiple parallel streams, with nesting and cross dependencies. They have seen the process from every angle—leader/member of the project under review, chair/member of the review team, neutral observer/learner, process improvement champion. This experience never ceases to bring new learning. There is simply no limit to how creative individuals in the process can affect the outcome.

There is always some room to revisit the process and make it a little bit more robust.

## 5 Re-thinking the Traditional Stage-Gate Process in Context of NDE 4.0

In light of the discussion above, the authors feel that the Stage-Gate process requires a critical reappraisal. This is useful for existing practitioners to improve the effectiveness, as well as for the new practitioners to get a strong start. The gate review should become an integral part of the business process. Particular attention should be paid to the workability of the tools and techniques for task integration and digital technology transfer across organizational lines. When implementing a new gate review procedure, build on existing tools and systems whenever possible. If possible, the new gate review process should be consistent with established project review procedures and management practices within an organization.

Overall project success depends on cross-functional integration via teamwork. Each task team should clearly understand the transfer mechanism for their work, be encouraged to seek out cooperation, and to check out early feasibility and integration. At times it is important to include into these interfaces support organizations such as purchasing, product assurance and legal services, as well as outside contractors and suppliers, especially if there are work interdependencies or issues affecting the project integration.

When introducing the stage-gate procedure, project leaders should anticipate anxieties and conflicts among their team members. These negative biases come from uncertainties associated with the new working conditions and requirements. They range from personal discomfort with skill requirements to anxieties over the impact of the new tool on the work processes and personal performance evaluations. This requires open communication and leadership trust.

### 5.1 Pilot Process

When revising or introducing a new management tool/process, such as a stage-gate, try it first with a small project and with an experienced, high-performing team, and even an external experienced review chair. Asking such a team to test, evaluate and fine-tune the gate review process for the company is often seen an honor and professional challenge. Further it usually starts the implementation with a positive attitude and creates an environment of open communications and candor.

Once proven, it should be documented and made a standard practice. Provisions must be made for updating and fine-tuning the stage-gate process on an ongoing basis to sustain relevancy. Like any other management tool,

stage-gate processes require top-down support to succeed. Managers can influence the attitude and commitment of their people toward the gate-review as a project control tool by their own actions. Concern for the project team members, assistance with the use of the tool, and enthusiasm for the project and its administrative support systems, can foster a climate of high motivation, involvement with the project and its management, open communications, and willingness to cooperate with the new requirements and use them effectively.

## 5.2 Recommended Practices

One possibility to avoid subjectivity and bias is that the success and failure criteria of a project is custom tailored, formalized, and held inviolate in the custody of some entity, similar to an IRB—Independent Review Board, which is the custodian and approver of all experiments and research programs and has a special focus on ethics, and adverse externalities. Data from the project is evaluated against those metrics at each gate review and the same is used to debias the decision process. The gate committee's human instinct to go against the data should be documented and analyzed to "calibrate the gut" once the final results are in. A caveat and key step must be to pivot hard on the criteria to correct misestimations.

It is easier to refine project-return estimations as a project nears launch. The unfortunate reality at many firms, is that near launch, attention shifts to delivery—and few like to disrupt execution. As a result, project managers often do not feel the need to bother with updating business cases with the latest insights. Even if it is late in the game, discontinuation remains hugely important, considering that most projects consume the majority of their development resources in those later stages, as things move towards mass production. Refer the sunk cost fallacy above. A single late-stage project can prevent dozens of alternative early-stage ideas from being funded. Failing to update business cases near launch, and thus missing signals of failure, ends up disproportionately expensive. Once again, a senior level person, not emotionally attached to the project should seek clarifications, and be willing to kill the project. Ronald Klingebiel [16] a professor of strategy at the Frankfurt School of Finance and Management in Germany, recommends having the role of business case sleuth. Such detectives could go after changes to business case assumptions when others have lost interest in evaluation and focus on getting across the finish line. Independent sleuths allow decision makers to build on new information about technological advancements, customer preferences, competitors' moves, or other factors with bearing on project business cases when these have the greatest resource implications. Averting one expensive fail stands

to more than pay for the extra business-case detective on your team.

The heightened attention to bad projects would be better placed on more promising alternatives. There is also the question of how much better a flagging business case can become, even if you look at it long and hard. Such attentional inertia can be reduced by minimizing the scope for interpretation and discussion. Setting clear discontinuation criteria beforehand ensures swifter, more automatic responses, preserving stage-gate decision makers' emotional energy for worthier pursuits.

If needed, split the review teams into two and play it like a mock court with defense and prosecutor to bring out hidden nuggets.

Finally, do not let decision paralysis set in when performance lags. Selective project progression is key in industries where investment occurs prior to knowing, and where learning during development determines the chances of success.

## 6 Case Study: Stage-Gate Process for Human-AI Collaborative Research and Reporting

We as humans, are beginning to see increasing role of domain specific and focused AI, also known as narrow AI in engineering design, manufacturing, supply chain, and decision making along the value stream. We regularly put the papers we write through tools like ProWritingAid [17]. These intellectual machines are now moving on from assistance to augmentation to becoming more like a partner.

### 6.1 Collaborative Co-author

The human authors recently gained access to beta release of the '*Generative Pre-trained Transformer 3*' (GPT-3), an autoregressive language model that uses deep learning to produce human-like text. It is the third-generation language prediction model in the GPT-n series created by OpenAI, a San Francisco-based artificial intelligence research laboratory. Authors also discovered that people have started using it to generate content [18–20]. GPT-3 is very powerful, albeit with some striking limitations as well [21]. The power comes from (a) 175 billion parameters, (b) training over a large portion of web pages from the internet, a giant collection of books, and all of Wikipedia and (c) tasks capability that include text classification (i.e. sentiment analysis), Question answering, Text generation, Text summarization, Named-entity recognition and Language translation. The limitations include (a) lack of long-term memory, (b) Lack of interpretability, (c) Limited input size, (d) Slow inference time, and (e) Does suffer from some bias, already.

## 6.2 Collaborative Research Methodology

The GPT-3 provides a robust REST API to access its endpoints. More specifically, authors collaborated through its completion endpoint using Python bindings to generate and explore the topics pertinent to this research. The modus operandi was surprisingly similar to collaborating with a savant with low domain knowledge that needed humans to be more explicit in their questions. The typical workflow was as follows. The human authors asked the AI a very specific question, with a very small output size. This step had to be performed at least a couple of times to generate an output that matched the human authors' intent. Since the AI is stochastic, a temperature setting corresponding to high creativity needed to be used to generate novel responses each time.

An example prompt is as shown in Fig. 4a. Once the output matched the intent, the new output was appended to the previous input (in the so called 'echo' mode), the AI was asked to generate the full output as desired. Authors noted that setting the frequency penalty and presence penalty parameters high led to promising results. In the example in Fig. 4b, the list was produced by running the first step. Once the list had some entries, the first bias was described, and GPT-3 was let free to do the rest.

This language model is an example of a "few shot" or a "zero shot" model, i.e., there is no gradient flow involved when the model seems to understand the context from a couple of examples, or from just an instruction. However, if the output is sufficiently involved, this can be fine-tuned with just a couple of hundred of pairs of inputs and expected

```
import os
import openai

openai.api_key = os.getenv("OPENAI_API_KEY")

response = openai.Completion.create(
    engine="davinci",
    prompt='Human Factors stage gate reviews in NDE 4.0 Investment Decisions\n
    Some of the human factors that can potentially affect the effectiveness
    of stage-gate reviews in Non-Destructive Evaluations 4.0 are:\n
    1.',
    temperature=0.7,
    max_tokens=64,
    top_p=1,
    frequency_penalty=0.8,
    presence_penalty=0.8
)
```

a

```
import os
import openai

openai.api_key = os.getenv("OPENAI_API_KEY")
## The list was produced by running the first step. Once the list had some entries, we described the first bias and let the AI do the rest
response = openai.Completion.create(
    engine="davinci",
    prompt='Question: Which human factors affect the outcome of a gate review in the phase gate process?\n
    Answer: 1. Bias. 2. Groupthink. 3. Stereotyping. 4. Recency. 5. Availability. 6. Representativeness.
    7. Anchoring and adjustment. 8. Confirmation bias. 9. Hindsight bias. 10. Probability neglect.
    11. Framing effects. 12. Optimism. 13. Overconfidence. 14. Unwarranted certainty. 15. Self-serving bias.
    16. Egocentric bias. 17. Illusion of control. 18. Illusion of validity. 19. Illusion 20. Backfire effect.
    21. Outcome bias. 22. Projection bias. 23. Overconfidence effect. 24. Attraction effect.
    25. Illusion of external agency. 26. Illusion of transparency.
    \n\n
    1. Bias affects gate review performance by influencing the interpretation of data and the selection
    of criteria for success and failure.\n\n
    2. Groupthink",
    temperature=0.7,
    max_tokens=201,
    top_p=1,
    frequency_penalty=0.8,
    presence_penalty=0.7
)
```

b

**Fig. 4** a Initial prompt to help GPT-3 understand the intent of the question. b An example of content generation based on cherry-picked output from the previous step



outputs. The fine-tuned model can then generate complex output such as test reports and documentation from raw data. The procedure can be extended by using a data cascade strategy to pair it with speech-to-text and similar technologies for zero touch operation.

### 6.3 Stage-Gates for the Collaborative Project

To collaborate with GPT-3 in context of NDE 4.0, the authors took it up as an innovation project—which means putting it through stages and gates. They kept it simple as

#### Human-Machine(GPT-3) Collaborative Research and Reporting Paper

**Stages:** Unstructured exploration and experimentation with AI Use in research and reporting for NDE 4.0 Initiative,

**Gates:** Strict assessment of content quality and relevance to NDE 4.0, at par with International Journal Standards



Fig. 5 Design of the stage gate process for creating this manuscript

Fig. 6 Details of the stage gate process for creating this manuscript

Stage-1	GPT-3 Capabilities Review & Assessment	<b>Actions:</b> Question answering, Text generation, Text summarization, and Ideation. <b>Result:</b> Surprisingly good with lots of useful information, and nothing wrong.
Gate-1	Can GPT-3 Add value?	<del>No, Not Likely, Don't know</del> , Likely, Yes
Stage-2	Content with AI-Augmentation	<b>Actions:</b> Create an outline of the essay. <b>Result:</b> Acceptable Outline, we made some changes. <b>Actions:</b> Find best practices in Stage-Gate process. <b>Result:</b> Acceptable list, could not match our experience. <b>Actions:</b> Identify human factors effecting gate review. <b>Result:</b> Extensive list with some valuable new items <b>Overall:</b> It validated what we had, added to our draft content, suggested ideas and summarized content.
Gate-2	Is the content relevant?	<del>No, Not Likely, Don't know</del> , Likely, Yes
Stage-3	Finalize Manuscript and perform AI Check	<b>Actions:</b> Assistance with abstract generation <b>Result:</b> The abstract made sense, with couple of iterations. It has been used with additional paragraphs to set the context, and novelty of using AI as a co-author
Gate-3	Does GPT-3 (AI) deserve to be a co-author?	Contributions from GPT-3 in this manuscript are at par with an intelligent graduate student or a junior colleague. It has made the manuscript more comprehensive and cohesive.  Recommend that GPT-3 be treated as a co-author, just like we would have done with an intelligent co-worker. This is also being honest to the reader community.
Stage-4	Publishers' assessment	<b>Actions:</b> Review of policy, procedures, and legal implications. <b>Result:</b> In process
Gate-4	Final Decision	Accept GPT-3 as a co-author of an AI augmented manuscript. or Move it to acknowledgement section as a tool used.

**Fig. 7** GPT-3 input and output (incrementally generated, 64 tokens at a time) to create an abstract for creating this manuscript

**INPUT:** Generate the abstract of the paper with the title: "Human Factors in NDE 4.0 Investment Decisions". The topics to touch include What is a Stage-Gate Process and its applications to NDE 4.0, Model for NDE 4.0 Initiatives, Implications of various outcomes of gate reviews in NDE 4.0, What helps a Stage-Gate Process be Successful?, Success Criterion, Review Team, Stage Execution, Gate Execution, What Makes a Stage-Gate Process Interesting?, Applicability, Success Metrics, Cognitive Human Biases that affect Gate review outcomes, Bricklayer's Fallacy, Stereotyping, Recency Bias and Availability Heuristics, Delicate Engagement, HiPPO Effect, Personal Insecurity, Anchoring and Adjustment, Framing Effects, Confirmation Bias, Probability Neglect, Sunk-cost Fallacy, Social Influence and Compliance, Other Human Biases, De-biasing our interpretation of Biases, Word of Caution, Re-thinking the Stage-Gate Process.

**SETTINGS:** engine="davinci-instruct-beta", temperature=0.7, max\_tokens=64, top\_p=1, frequency\_penalty=0.8, presence\_penalty=0.61

**OUTPUT:** Stage-gate process is a project management technique that has been used to manage major engineering projects for over 50 years. The stage gate process emphasizes quality and value and provides an orderly framework for decision making with multiple projects carried out concurrently. This paper will examine the potential application of this process to NDE 4.0 investments. This paper will also explore the factors that make a stage-gate process successful, and how biases affect the review outcomes of gate reviews. It explores which biases are problematic and how we can de-bias our own interpretation of them through conscious awareness and re-thinking the Stage-Gate Process. It also discusses how to create an environment where biases are minimized or eliminated.

shown in Fig. 5, with detailed actions and gate outcomes presented in Fig. 6.

Outcome of the Gate-3 is merged within content of this manuscript. Once finalized, authors used the table of contents to create the abstract. Figure 7 shows the original abstract created by GPT-3. It made sense after a couple of incremental iterations. GPT-3 requires certain parameters and settings, just like any human research assistant would need instructions and boundaries to have a meaningful outcome. DaVinci Instruct is the engine most suitable for following instructions, rather than using them as text prompts. Parameters were empirically chosen for suitable results. Temperature is the parameter corresponding to the "sampling type". 0.7 is the default for new content generation. Lower temperatures give more predictable results with less "creativity"; Top\_p = 1 is another setting to control sampling using "nucleus sampling" which was kept at the default value of 1; Frequency penalty and presence penalty nudge the AI to not repeat words and topics in the output respectively.

The abstract generated by GPT-3 has been used at the beginning of the manuscript with additional paragraphs to set the context and define the novelty of using AI as a co-author.

For NDE 4.0 developers, GPT-3 is a potential partner for intelligence augmentation in interpreting information and creating reports, examples applications being around keyword extraction, summarization, semantic information retrieval.

The first round at Gate-4 was a conditional go. Human authors received reviewer feedback and incorporated many additions to the manuscript. The final outcome of the Gate-4 will not be known until the paper is published. For the reader, it should be evident from the list of authors.

#### 6.4 Human Author Experience

There were human factors during the stage-gates of case study execution as well. Human authors were skeptical about the capability in the beginning and amazed later on. During

the final review gate and revision stage, despite some experience, there was a noticeable tendency to treat the machine co-author as a tool/special entity, with humanly desire to identify parts of the treatise written by GPT3 differently. It took special effort for human authors to explicitly check and debias themselves to not treat GPT3 any differently. Authors conversations with several peers were met with varied levels of curiosity and suspicion. Readers of the paper might feel the same way and may need some time and adjustment to fully absorb the process and impact. Considering the importance and relevance of the topic, the journal editors accelerated the review process, but were very conscious of relevance and standard of the content. There was no compromise on quality and acceptability criterion.

The regulatory and legal concerns with publication, if any, are still in the future by the time this text is written. Regardless of the outcome of Gate-4, in the spirit of innovation, where you explore and accept the outcome as a learning, this collaboration has been an eye-opening and rewarding experience. It provided a convincing evidence of a powerful human-machine coworking at linguistic and cognitive levels, one which is closer than we think, and more powerful than we conceive. The speed at which the research can be conducted is exponentially faster. This is just like digitalization of third revolution which had offered an order of magnitude faster access to information as compared to physical libraries of indexed books. Ability of AI to connect pieces of information into a coherent story can take creativity to a whole new level.

From NDE 4.0 perspective, it offers unimaginable opportunities in developing and writing NDT procedures, creating maintenance plans, creating inspection reports, adapting maintenance plans to inspection outcomes. GPT-3 is now embarked on extending this capability to images. And that will open up another new dimension in NDE 4.0 heavily dependent upon image processing. When combined with voice activated devices, it can soon get to a point where humans may wonder if their assistant is really a machine.

## 7 Closing Remarks

The effective management of stage-gate process and review meetings involves a whole spectrum of critical factors: clear direction and guidance; ability to plan and elicit commitments; communication skills; dealing effectively with managers and support personnel across functional lines often with little or no formal authority; information-processing skills, the ability to collect and filter relevant data valid for decision making in a dynamic environment; and ability to integrate individual demands, requirements, and limitations into decisions that benefit the overall project. It further involves the project leader's ability to resolve intergroup

conflicts and to build multifunctional teams. Several practices described in this paper have been derived from the broader context of this field study to help both NDE 4.0 project leaders and managers to understand the complex interaction of organizational and behaviors involved in innovations.

The stage-gate process is designed to minimize cost and risk of innovation project through synergy and alignment of expectations. A well-defined and well-controlled process for development and adoption of NDE 4.0 philosophy:

- (a) Identifies the makeup of a digital-physical expert review team and lays out the decision criteria upfront in context of the NDE 4.0 use case.
- (b) Provides a forum and timing to discuss and approve any scope changes,
- (c) Uses a common language across NDE practitioners and data scientists,
- (d) Maintains a nonthreatening atmosphere,
- (e) Clarifies and adapts the roles & responsibilities during execution,
- (f) Identifies intellectual property and other business protection needs,
- (g) Facilitates informed decision making for the continuation of the NDE 4.0 project based on the availability of resources, business use case, and risk analysis.

However, for the process to deliver to its promise effectively, the organization must deliberately create an environment of healthy conflict at the Gate reviews. The review team must be-

- (1) Competent to make the right decisions, despite the emotional attachment with concept,
- (2) Empowered to judge and stop or redirect a project, despite business pressures, and
- (3) Objective to minimize bias towards NDE 3.0 and emotional interference.

Good gate reviews are a work of art and science. During a revolution, they also require a bit of an exploratory streak with self-confidence.

The collaborative work between two humans and an artificial intelligence partner GPT-3 has opened a new chapter in research and reporting. The authors expect more such studies in the future. This speeds up research, enhances the breadth of input, and can reduce bias when handled carefully. The contributions from GPT-3 in this manuscript are at par with an intelligent graduate student or a junior colleague. It has made the outcome more comprehensive and cohesive. However, it cannot work independently since it is not yet mature enough to have subjective experiences.

Partnership with GPT-3 shows promise as a value driver [8] in NDE 4.0 by serving the role of a Digital Twin assistant

to humans in the cyber-physical loops of an NDE ecosystem [8, 22]. It is worth investing.

**Acknowledgements** Authors acknowledge the willingness of OpenAI to provide access to the GPT-3 Beta version. They are also grateful to Dr. Johannes Vrana of Vrana GmbH and Dr. Anukram Mishra of Genus Power Infrastructures for their continued support of this collaboration.

## References

- Vrana, J., Singh, R.: Digitization, digitalization, and digital transformation. In: Meyendorf, N., Ida, N., Singh, R., Vrana, J. (eds.) *Handbook of Nondestructive Evaluation 4.0*. Springer, Cham (2021)
- ISO 56002:2019: Innovation Management Standard—Guidance
- Singh, R.: Purpose and pursuit of NDE 4.0. *Spec. Issue Mater. Eval.* **78**, 785–793 (2020)
- Singh, R.: *Inspiring Next Innovation Mindset—Why Not ... Start with the Why Not?* Outskirts Press, Book (2020)
- Vrana, J., Singh, R.: Introduction to NDE 4.0. In: Meyendorf, N., Ida, N., Singh, R., Vrana, J. (eds.) *Handbook of Nondestructive Evaluation 4.0*. Springer, Cham (2021)
- Singh, R.: *Inspiring Next Innovation Value Chain—How About ... Best in the World?* Outskirts Press, Book (2020)
- Vrana, J., Singh, R.: NDE 4.0—a design thinking perspective. *J. Nondestruct. Eval.* **40**, 8 (2021). <https://doi.org/10.1007/s10921-020-00735-9>
- Vrana, J., Singh, R.: Cyber-physical loops as drivers of value creation in NDE 4.0. *J. Nondestruct. Eval.* **40**, 61 (2021). <https://doi.org/10.1007/s10921-021-00793-7>
- Mihaly, H.: From NASA to EU: the evolution of the TRL scale in Public Sector Innovation. *Innov. J.* **22**, 1–23 (2017)
- U.S. Government Accountability Office: *Best Practices—Better Management of Technology Development Can Improve Weapon System Outcomes*. NSIAD-99-162 (July 1999)
- U.S. Government Accountability Office: *Best Practices—Capturing Design and Manufacturing Knowledge Early Improves Acquisition Outcomes*. GAO-02-701 (July 2002)
- Singh, R.: NDE 4.0—the next revolution in nondestructive testing and evaluation: what and how? *Mater. Eval.* **77**, 45–50 (2019)
- Thamhain, H.: Applying stage-gate processes in concurrent engineering. *Wescon/96*, 1996, pp. 2–7, (Oct 1996). <https://doi.org/10.1109/WESCON.1996.553955>. <https://ieeexplore.ieee.org/abstract/document/553955>
- Sethi, R., Iqbal, Z.: Stage-gate controls, learning failure, and adverse effect on novel new products. *J. Mark.* **72**(1), 118–134 (2008)
- Pohl, R.F.: *Cognitive illusions: a handbook on fallacies and biases in thinking, judgement and memory*. Psychology Press. ISBN 978-1-84169-351-4 (2004)
- Klingebiel, R.: Research: how to get better at killing bad projects. <https://hbr.org/2021/04/research-how-to-get-better-at-killing-bad-projects> (April 02, 2021)
- TNW News: ProWritingAid sets AI loose to improve your writing like a trusted editor. <https://thenextweb.com/news/prowritingaid-sets-ai-loose-to-improve-your-writing-like-a-trusted-editor> (July 01, 2021)
- GPT-3: A robot wrote this entire article. Are you scared yet, human? <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3> (Sept 08, 2020)
- Porr, L.: Can AI and GPT-3 replace authors? Writer on the side. <https://podcastnotes.org/writer-on-the-side/can-ai-and-gpt-3-replace-authors/> (Oct 21, 2020)
- Aalho, J.: I wrote a book with GPT-3 AI in 24 hours—and got it published. <https://medium.com/swlh/i-wrote-a-book-with-gpt-3-ai-in-24-hours-and-got-it-published-93cf3c96f120> (June 12, 2021)
- Mavuduru, A.: What is GPT-3 and why is it so powerful? <https://towardsdatascience.com/what-is-gpt-3-and-why-is-it-so-powerful-21ea1ba59811> (Feb 17, 2021)
- Vrana, J.: The core of the fourth revolution: industrial internet of things, digital twin, and cyber-physical loops. *J. Nondestruct. Eval.* **40**, 46 (2021). <https://doi.org/10.1007/s10921-021-00777-7>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.