契約締結におけるAI品質ハンドブック

2021年7月

一般社団法人日本ディープラーニング協会
「契約締結におけるAI品質保証の在り方」研究会

Japan
Deep Learning
Association

はじめに

昨今、ディープラーニング技術の進展と認知の広がりと共に、これらの技術を本格的に製品・サービスやビジネスプロセスに導入し、実用化しようとする動きが活発化しています。AI システムの本格導入を進めたい企業にとって、優れた AI 技術と高い機動性を持つ AI 開発スタートアップへの期待は高まっていると言えます。その一方で、AI に対する特質性の捉え方の違いや過度な期待等により起こる契約上の課題も顕著になってきていることが覗えます。

特に「品質」に関して、ディープラーニング等の機械学習による手法は、本番導入後も学習を継続し続けるというその性質から、開発契約時点に従来型のソフトウェア開発時と同様の概念で品質を捉えて保証することは困難です。しかし、実際は、ある一定の性能保証を条件にすることをめぐり交渉が難航したり、導入後に思ったような精度が出なかったりすることが契約上の問題となる等、プロジェクトの円滑な進行を妨げる要因にもなっています。また性能以外の品質として挙げられる「説明可能性」や「公平性」等の AI の信頼性に係る要素は、プロダクトの活用領域によっては今後ますます重要になると考えられ、十分な協議をしないでプロジェクトを進めることは、後々の法的トラブルにつながるリスクがあります。

以上の現状認識から、本研究会では、AI 開発時の合意形成において留意する点をまとめ実用的に使えるようハンドブックを作成しました。委託側と受託側の双方が、品質について共通の認識をもって必要な協議を行えるように、既存の関連ガイドライン、実際の事例等を参考にしつつ、特に開発時点における品質の捉え方と確認のポイントを整理しています。また、AI 品質は、プロダクト種類や領域によって重要な論点も異なります。そのため研究会では、実際の開発事例を踏まえて、プロダクト種類毎に論点を示すことを試みました。事例ヒアリングには、研究会メンバーを初め多くの JDLA 会員企業や有識者会員の皆様にご協力をいただきました。改めて感謝申し上げます。

本ハンドブックをコミュニケーションツールとして活用することで、円滑かつ効果的な案件進行につながれば幸いです。発注側として主に想定できる大企業側においては、AI 品質特性への理解が深まり、プロジェクト遂行への不安が解消し、主体的かつ合理的な評価項目の検討が進むこと等が考えられます。受託側として主に想定するスタートアップ企業では、発注元の期待や要件の的確な把握が進むことでより効果的な提案ができること、また契約交渉にかかる工数の削減といった効果も想定しています。
AI 品質の向上には、導入後の検証等、より継続的なプロセスが必要と言え、発注企業と開発業者間が十分に協議し、共に創る意識が求められます。開発時の AI 品質について、これまで以上に質の高い議論が進むことで、広く産業における AI プロダクト・サービスの品質向上や信頼性の担保へとつながることを期待します。

（一社）日本ディープラーニング協会
「契約締結における AI 品質保証の在り方」研究会
座長　南野 充則

目次

1. 本ハンドブックのスコープと用語の定義

1-1 本ハンドブックのスコープ

1-1-1 既存の品質ガイドラインとの関係

AI モデルの品質について分析・検討したガイドラインとしては機械学習品質マネジメントガイドライン（産総研）と AI プロダクト品質保証ガイドライン（QA4AI）がある。もっとも、それらのガイドラインはあくまで、AI モデルを組み込んだ製品・サービスが最終システムの利用者に対して提供される際の品質（利用時品質）について分析・検討したものである。

一方、本ハンドブックは、ディープラーニング技術を使った AI ソフトウェア開発を行うスタートアップ企業と、自社のビジネスへの導入のためにそれらの開発を委託しようとする企業が契約をする場合に、当該利用時品質を実現するためにどのような点に留意して交渉をすべきか、という点に焦点を当てて整理したものである。

1-1-2 本ハンドブックの対象となる契約のフェーズ

AI 開発では、アセスメントや PoC フェーズといった事前の検証プロセスを経る場合が多いが、ここでは、それらの検証フェーズ後、実導入に向けて学習済みモデルの生成を行う段階における開発契約を想定する。

また AI 開発における特徴的な論点を示すという趣旨から、契約の対象としては、開発する AI モデル部分のみを捉える。実際の運用に向けて実装が必要となり得る AI モデルを含むシステム全体に関しては、AI モデルの品質を協議する際に特に留意する観点がある場合に触れるのみに留める。

1-1-3 本ハンドブックの対象とする品質特性の種類

本ハンドブックでは、既存の関連ガイドライン（機械学習品質マネジメントガイドライン（産総研）、AI プロダクト品質保証ガイドライン（QA4AI））を参照し、AI 開発に重要と考えられる品質特性を 5 項目（「性能（有用性・リスク回避性）」「公平性」「頑健性」「説明可能性」「セキュリティ」）ついて示す。開発契約交渉時の論点としてよく挙げられる性能の観点に加え、AI の信頼性向上の観点から今後重要になると思われる観点も含めて対象としている。

1-2 用語の定義

（1）「AI」＝ ディープラーニングの手法等を用いた帰納的に開発する機械学習技術

（2）「品質保証」＝一定の性能を持ったモデルの作成義務を受託者が負う場合だけでなく、そのような義務を負うわけではないが一定の性能を持ったモデルを作成できない場合には代金を受領できないという場合も保証に含めるものとする。

（3） 評価指標[1]：

「正解率」＝ 全データ中、どれだけ予測が当たったかの割合

「適合率」＝ 予測が正の中で、実際に正であったものの割合

「再現率」＝ 実際に正であるものの中で、正だと予測できた割合

「F 値」 ＝ 適合率と再現率の調和平均

[1]（一社）日本ディープラーニング協会、『ディープラーニング G 検定公式テキスト』p.143

2. AI 開発契約における品質保証の在り方

2-1 従来型のソフトウェア開発契約における「品質」の位置づけ

従来のソフトウェア開発契約においても、ソフトウェアの「品質」は重要な交渉ポイント及び合意対象となる。最終成果物が当該品質を満たしていなかった場合には、契約不適合責任に基づき、成果物の無償補修、代金減額、損害賠償、解除などが問題となるためである。

もっとも、裁判においては、かかる「品質」の合意がなされているか自体が争点となることが多い。

要件定義書などの仕様書に「品質」が明記されていればそれにより判断されるが、そのような記載が無い場合、裁判所は、契約書、検収書、議事録等の各種書類によって当事者の意思を推測していくことになる。また明示的に当事者の意思を確定できない場合であっても「当然備えるべき品質」（たとえば最低限のセキュリティなど）については品質に関する合意がなされていると認定されることもある。

2-2 AI モデル開発契約における「品質」交渉の難しさ

一方、AI モデル開発契約においては、開発が帰納的に行われるという技術的特性から、「実環境における当該ソフトウェアの利用時に一定の品質を有する旨を契約締結時に合意すること、すなわち一定の利用時品質を合意することがそもそも困難である。

また、AI モデル開発の場合、技術的な要素に加え、そもそもユーザが成果物に関するどのような「品質」をどのレベルで希望しているのかが必ずしも明確でない場合も多い。

そこで、2-3 ではそれらの AI モデル開発契約の特質を踏まえた対応方法のポイントについて解説する。

2-3 AI 開発の特殊性を踏まえた対応方法

2-3-1　利用時品質と開発時品質の区別

まず AI モデルの開発契約締結交渉において「利用時品質」と「開発時品質」を明確に区別し、ユーザとコミュニケーションを図ることが重要である。

ここで「利用時品質」とは、開発が完了した AI モデル（同モデルを構成要素とするシステム全体も含む）を実環境に投入した場合における品質を言い、「開発時品質」とは、開発が完了した時点における品質（言い換えれば、開発が完了した AI モデルにテスト用データを入力した際の出力の品質）をいう。

ユーザが「品質を保証して欲しい」と主張する際の「品質」とは「利用時品質」のことを意味することがほとんどである。ユーザにとっては実環境における品質こそが関心事だからである。

そして、従来型のソフトウェアにおいては演繹的に開発されることから「利用時品質」の保証を行うことも可能だが、帰納的に開発される AI モデルにおいては、開発者が保証できるのはあくまで「開発時品質」であって、「利用時品質」を保証することは原理的に不可能である。実環境においては学習データセット（訓練用データ、テスト用データ、バリデーション用データからなる）とは異なる偏りを持つデータが入力されるためである。

その点についての開発者・ユーザ間のコミュニケーションが曖昧だと、たとえば、契約締結時に「一定の精度を保証する」という合意をした場合の「精度」について「利用時品質」と「開発時品質」のどちらなのかについての認識の相違が生じ大きなトラブルとなる。

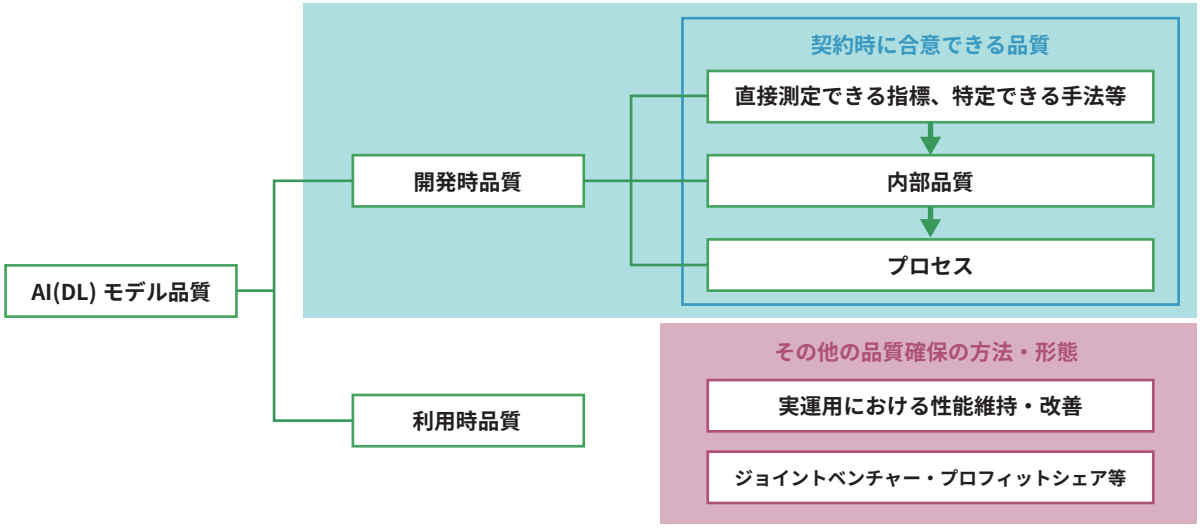


図１：AI（DL）モデル品質における契約の考え方

2-3-2　ユーザは、どのような「品質」について、どのレベルの保証を求めているのか

（1）「品質」の種類（≒品質特性）

工場における異常検知 AI の開発においてユーザの希望が「100 パーセント人間が作業をしている現状と比較して、導入効果がある『いい感じの AI』を作ってほしい」だったとしよう。この場合における「いい感じ」つまりユーザが求めている品質は、通常 2 種類の意味が考えられる。

1 つは「リスク回避性」である。

これは工場における異常検知 AI の例でいえば「異常品を正常品と誤判断する誤判定率を極力小さくする」（検出漏れの防止）という品質を意味する。

もう 1 つは「有用性」である。

これは、同じく異常検知 AI の例でいえば「正常品を異常品と誤判断しないことにより、100 パーセント人間が作業をしている現状と比較して人間の作業時間が減少する」（誤検知の防止）という品質を意味する。

ユーザはかならずしも上記 2 種類の品質を区別して認識しているわけではないので、開発者としてはユーザが「いい感じの AI」として、どちらの（あるいは両方の）品質を重視しているのかを確認する必要がある。

以上は「工場における異常検知 AI の開発」においてユーザが求める品質として「リスク回避性」「有用性」について述べたが、もちろん、それ以外にも開発対象となる AI モデルの種類に応じて異なる品質が求められることがある。本ハンドブックでは、AI モデル開発にかかる品質の種類を5つの「品質特性」:「性能（有用性・リスク回避性）」「公平性」「頑健性」「説明可能性」「セキュリティ」に整理して進める。

（2）　求められる「品質」のレベル

さらに、「品質」についてはユーザによって要求するレベルが異なる。

たとえば、「リスク回避性」については「誤判定すれば人の死亡を含む重要な結果が生じる」かつ「人間が関与することなく AI モデルの判断要素がそのまま利用される」のであれば、非常に高いリスク回避性が必要となる一方で「誤判定したとしても人的な損害はなく、経済的な軽微な損害が生じるのみ」かつ「AI モデルの判断結果を元に人間が最終的に判断する」のであれば、低いリスク回避性でも問題がない。

また、「有用性」についても「当該製品・サービスが一定の性能指標（正答率・適合率・再現率など）を満たすことが、製品・サービスの運用上の必須または強い前提である場合」「一定の性能指標が製品・サービスの目的として特定されているが、リリースまでの日程スケジュールが重視される場合、または品質をモニタリングしながら試験運用の実施により漸次性能向上を行う運用が許される場合。」「性能指標が開発時点で特定されず、性能指標そのものを発見することが開発目的となる場合」など異なる要求レベルが存在する。

(3)　まとめ

すなわち「品質保証」の問題については、「品質の種類（≒品質特性）」×「求められる品質のレベル」に分解して交渉をすることが重要である。

たとえば、冒頭の事例（工場における異常検知 AI の開発）おいては以下のようなやりとりが考えられる。

> **ユーザ**：100パーセント人間が作業をしている現状と比較して、導入効果がある『いい感じの AI』を作ってほしい。
>
> **開発者**：「いい感じの AI」とは具体的にどのような意味でしょうか。今回、AI を利用して異常検知しようとしている部品については、「本来異常であるものを正常と判断してしまう」という検出漏れをなるべく少なくしたいのか、「本来正常であるものを異常と判断すること」を減らし、人間による二重チェックの手間を減らすという意味での効率性を重視したい（品質の種類）と言うことなのか、どちらでしょう。
>
> **ユーザ**：対象部品は非常に重要な部品であり、現状は 2 人の人間が全ての部品を目視にてダブルチェックしている。現状では再現率（検出漏れの少なさ）は100%なので、それと同等の再現率にして欲しい。
>
> **開発者**：AI モデルの技術的な特性から 100%の精度を出すことは難しい。もっとも、検出漏れがあったのでは致命的だろうから、適合率（誤検出の少なさ）を若干犠牲にしてもいいから再現率（検出漏れの少なさ）を優先させる（求められる品質のレベル）ということでよいか。そのような AI であっても、AI が異常と判断した製品のみを人間がダブルチェックするという体制をとれば、現状の「2 人の人間が全ての部品を目視で確認する」よりも人手はかなり少なくてよくなる。
>
> **ユーザ**：それでよい。

2-3-3「品質保証」とは何を「合意」するのか

「品質保証」の問題については、「品質の種類（≒品質特性）」×「求められる品質のレベル」に分解して交渉をすることが重要であると説明したが、ではその交渉の結果、当事者間で品質保証に関して何を合意するのか。

この点については品質特性ごとに異なるため、詳細は「3　品質特性ごとの対応について」を参照されたい。

2-3-4「開発契約時点において品質を合意すること」以外の品質確保の方法・形態

もっとも、「開発契約時点において品質を合意すること」自体が困難なこともありうる。結局、ここでの問題は「将来判明するかもしれない品質不足のリスク」「将来の運用において必要となるかもしれない品質改善のためのコスト」をどちらが負担し、「将来、AI モデルの品質が向上することによるリターン」をどちらが得るかを、AI モデルの特質上、契約締結時点では合意できないことにある。

以上を踏まえると、「開発契約時点において品質を合意すること」以外の品質確保の方法や形態として下記の方向性も考えられる。

（1） 開発後、実運用での性能維持・改善（例：追加学習等）を合意する。
利用時品質については契約締結時点で合意することは極めて困難・不可能であるため、契約時点において合意することはあきらめ、開発後実運用における性能維持・改善を合意する。

（2） プロフィットシェア、ジョイントベンチャー
実装後利用時品質が変化する（＝悪化する）リスク、利用時品質を改善することのコスト、利用時品質が改善されたことによって得られる利益を合理的に分配するために、プロフィットシェアやジョイントベンチャーを採用する。

3. 品質特性毎の対応について

ここでは、「性能（有用性・リスク回避性）」「公平性」「頑健性」「説明可能性」「セキュリティ」の 5 つの品質の観点から、開発契約時に合意する際の留意点について示す。

3-1 性能（有用性・リスク回避性）

（1） 性能（有用性・リスク回避性）とは

一定の指標を用いて計測される性能をいう[2]。「有用性」と「リスク回避性」の観点で示す。

有用性

AI モデル要素によって効果がもたらされる度合いを指す。（指標の例：正解率 90% 以上、適合率 80% 以上、F 値 70% 以上、等）

リスク回避性

安全性を追求するタイプの品質であり、AI モデル要素の誤判断によって悪影響（人的被害・経済的被害）を及ぼすリスクを回避・低減することを目的とするもの を指す[3]。（指標の例：誤判定率 5％以下、誤認識率 5％以下、過検出 4%、検出時間 1秒以内、等）

（2） 性能の合意方法に関する留意点

何を評価指標とするかはタスク内容、利用方法に依存する。また、契約当事者がどのような指標を重視するか、単一・複数の指標を採用するのか、案件毎に検討して協議することが重要となる。また、開発する AI が意思決定の自動化を目的とするものかどうかも重要な視点となる。最終的な意思決定に人間が関与する場合は、関与する人間の負荷を軽減する範囲でリスク回避性を考慮することになる。

有用性

「正解率 90%」のように指標を用いて合意することで保証は可能だが、実際のところ AI モデルの出力はデータの量および質に依存するため、契約段階では合意、数値保証を行うことは困難である。実務上は PoC を前提とした同等の精度を保証することが考えられるが、この場合も PoC と本番環境では通信速度やハードウェアに差が存在することに留意する必要がある。

リスク回避性

リスク（危害の発生確率及びその危害の度合いの組合せ）を想定した上で、リスクを引き起こす誤判定率、誤認識率を限りなく小さくする、決められた時間内に検出する、といった指標が考えられることから、具体的に誤判定率、誤認識率、検出時間というような指標を用いて合意することで、一応の品質保証をすることが可能である。

ただし、外部環境に依存するシステムにおいては、データの変動要因が多数存在する場合もある。この場合には、予期しない（未経験の）変動要因に対し、PoC や開発時にすべてを仕様化し、扱うべきデー

[2] SQuBOK 策定部会『ソフトウェア品質知識体系ガイド（第 3 版）－SQuBOK Guide V3－』（オーム社、2020 年）、p.285

[3] 国立研究開発法人産業技術総合研究所「機械学習品質マネジメントガイドライン第 1 版（2020 年 6 月 30 日）」（https://www.cpsec.aist.go.jp/achievements/aiqm/AIQM-Guideline-1.0.1.pdf）、p.11

タすべてを分類、検証することは不可能であり、保証は困難と言える。

※自動運転の場合（極めて高いリスク回避性）
自動運転で完全無人化を目指す場合は、極めて短い時間で対象物を的確に検知するという、物体検出率やスピードについて極めて高い安全性のレベルが求められる。
誤認識（運転車による回避操作が不可能。円滑な交通を阻害）と未認識（必要なシーンでブレーキをかけない。オーバーライドで回避可能）で影響が異なることを考慮しながら評価指標として設定することが重要になる。（指標の例：検出時間、閾値（誤認識による円滑な交通の阻害と事故発生時の重大性の調和がとれる適合率））。また、物体が何かを認識できていなくても、物体に接近するだけでアラートとブレーキをかける衝突被害軽減ブレーキのように、機械学習を利用したシステムとは独立した安全に関するシステムを利用するといった役割分担も想定される。

3-2 公平性

（1） 公平性とは

公平性とは、AI の出力が公平であるか、バイアスを有していないかの点に関する品質特性である。

（2） 公平性の合意方法に関する留意点

公平性については、「男女間で正解率の差異が 20% 以内」のように指標を用いて合意することで保証することが可能とはいえる。ただし、性能の場合と同じく AI モデルの出力は、学習用データに依存するため多くの場合では保証は難しい。

公平性を実現するには属性の特定が重要となる。機械学習とはデータの違いに応じて判定を行うものである以上、一定の属性に基づいて出力に差異を付けることは認められる。例えば、貸付判定の場合に年収を考慮するなどである。すなわち、許されるバイアスと許されないバイアスが存在し、個別のユースケース毎にこれらを判定する必要がある。

なお、バイアスの事例として著名な COMPAS に関する事案では、入力特徴とされていない人種が他の入力特徴から推測され人種によるバイアス[4]が指摘されていた。この点からも、入力特徴のみに着目するのではなく、入力特徴から推測される属性も含めて検討することやそもそも入力特徴を離れて検討することが必要となる。

3-3 頑健性

（1） 頑健性とは

[4] 刑事被告人の再犯リスクのスコアを予測する COMPAS では白人と黒人間の偽陽性率（再犯リスクが高い場合を陽性とする。つまり、真実は再犯リスクが高いのに誤って低いと予測した割合である。）の差異が批判されている。

ここでは、学習段階および運用段階における頑健性を指す。
学習段階における頑健性とは、外れ値やノイズ、過学習を避けて推定性能の高いモデルを実現することを意味する。運用段階における頑健性とは、外れ値やイレギュラーなことが起きた時でも推定性能が高い（汎化性能が高い）ことを意味する。

（2） 頑健性の合意方法に関する留意点

学習段階における頑健性

例えば、ノイズや外れ値が含まれる場合の学習方法（例えば、平均値は、外れ値に対して頑健ではない）のように、アルゴリズムの選択根拠やパラメータ、ハイパーパラメータの設定根拠を説明した上で合意することが考えられる。

また、ノイズ候補の洗い出しとノイズ候補によりどの程度学習済みモデルが劣化するか、データ量を変化させることで、どの程度、学習済みモデルの性能が変わるかを記録することを合意することが考えられる。これは、学習の進行に応じて推定性能が劣化した場合に、その性能劣化が許容可能な範囲か、影響範囲をきちんと把握できているか、そのデグレードが再現可能かを把握することを目的とする。

ただし、これらは開発過程の試行錯誤する中で根拠が説明できるようになることも考えられ、契約時点に合意することは困難である。その場合の合意の対象は、精度を確認するための適合率、再現率、F 値等の指標について、数値を合意することになる。

運用段階における頑健性

産業用システムにおける頑健性は、環境依存性により、工場などのクローズドなシステムであっても、データの変動要因が多数存在することから、予期しない（未経験の）変動要因に対し、PoC や開発時にすべてを仕様化し、扱うべきデータすべてを用意して検証することは不可能である。

そのため、頑健性の評価に必要な網羅性を、収集したデータのみで数理的に担保することが難しいため、事前に合意した評価指標や AUROC（ROC 曲線における AUC）といったモデルのよさを示す指標によって汎化性能を測ることになる。また、精度測定の方法等についても合意をすることが考えられる。

3-4 説明可能性

（1） 説明可能性とは

説明可能性とは、明確な定義が存在しないが、ここではモデルの有する判定根拠を示す能力を言うものとする。

（2） 説明可能性の合意方法に関する留意点

説明可能性についてはこれを直接的に数値化して測定することが難しく、ユースケースを分析して仮に説明可能性が必要である場合、必要とされる説明可能性の内容などを分析のうえ、採用するアルゴリズムを決定木とする、Grad-CAM などの説明可能性を付与する技術を実装するといった対応を行うことで説明可能性を付与することになろう。このため、要件定義書などで採用するアルゴリズムや実装する説明可能性に関する技術を定めることで、説明可能性に関する合意を行うことになる。

3-5 セキュリティ

（1） セキュリティとは

AI に関するセキュリティとして、ここでは通常のシステムに見られない攻撃方法[5]を想定する。

（2） セキュリティの合意方法に関する留意点

システムのセキュリティは度合いを直接数値化して測定することが難しく、通常のシステムでは想定される攻撃方法を検討し、リスクの度合いなどの必要に応じて対応する方法を要件定義書にて定めてゆく。例えば、通信の暗号化という方法を要件定義書に定めるなどである。AI の場合も、これと同じように想定される攻撃方法を検討の上、必要に応じて対応する方法を要件定義書などにて定めてゆくことになる。つまり、セキュリティにおいても説明可能性と同じく、要件定義書などで実装するセキュリティ技術を定めることでセキュリティに関する合意を行うことが考えられる。

ただし、AI のセキュリティに関して実務レベルでの対応例はまだほとんど見られないのが現状である。重要な要素ではあるが、契約時の落とし込みについては研究等の動向を踏まえて今後の課題といえる。

【指標を合意する場合に必要な確認点】
ある指標を用いて合意する際、以下の項目についても定義することが必要

① 測定方法の設定
- ホールドアウト法：テスト用データを別に取っておく
- クロスバリデーション法：複数の学習を実施して最終的に全データを学習用に利用する

② 測定主体の設定
- ユーザ
- ベンダ
- 第三者機関

テスト用データを用いて性能を測定する際、推論コードや精度評価指標のコードが必要になるため、AI に関する知見やプログラミング能力が要求される。よってベンダが測定主体となることが多い。テスト用データを学習に用いることで生じる過学習のリスクに注意。
場合によっては、発注者側が適切なテスト用データを第三者機関に提供し、テストを行うことも考えられる。

---

[5] 具体的には、データ汚染攻撃、回避攻撃、オラクル攻撃等が想定される。攻撃種類の詳細に関する参照先としては、https://www.nccoe.nist.gov/projects/building-blocks/artificial-intelligence-adversarial-machine-learning（アメリカ国立標準技術研究所（NIST））や、https://www.imes.boj.or.jp/research/abstracts/japanese/20-J-20.html（菅和聖「機械学習システムの脆弱性とセキュリティ・リスク：「障害モード」による分類と今後へのインプリケーション」日本銀行金融研究所、Discussion Paper No. 2020-J-20）等がある。

③　テスト用データの設定

・開発時に存在するデータ（の一部）

・運用開始後のデータ　※実務上ほとんど行われていない

精度測定用のテスト用データの内容を予め設定、合意する。
開発時点で未知のデータに対してモデルがどのような挙動を示すかは予測できないため、保証を行うとしても開発時のデータをテスト用データとすることが一般的である。
なお、運用開始後の性能については保守や再学習で精度向上を行うことができる。
また、テスト用データの構成についても適切な設計が必要となる（第4章に詳述）。なお、テスト用データについては、数理的な多様性だけではなく、意味的な多様性や、社会的・文化的な多様性についても考慮することが望ましい。

4．開発段階の検討項目の詳細化と保証（合意）の可否

開発段階で品質を確保する対象として想定できる項目の例を示し、保証（合意）可能なもの、保証（合意）できないものについて示す。

本ハンドブック作成時点において、リスト項目について合意し開発を進めている案件は少ないものの、将来的にはリスト項目にかかる品質保証が求められうること、紛争予防のために可能な範囲で合意しておくことが望ましいことから、詳細にリスト化している。

リスト項目作成においては以下を前提としている。

- 3 に記載の対応方法と重複する項目はハイライトにしている。
- 3 において、実際には保証が困難とされるもの（評価指標の確定等）についても、保証することができることを前提にリスト化している。
- リスト項目すべてを遵守しなければならないわけではなく、プロダクトの開発において要求される品質レベルに応じて、必要かつ合意可能な項目を読者が適宜選択することを想定している。
- リスト項目は一例であり、列挙された項目以外の事項について合意をすることを否定するものではない。
- 品質レベルとリスト項目の関係について、品質レベルに応じて合意すべき項目は増えるが、各項目について保証の可否が変わるわけではないため、品質レベルを考慮せず、単に保証できるか否かという視点で項目を列挙した。
- リスト項目の作成にあたっては、機械学習品質マネジメントガイドライン（第1版）、AI プロダクト品質保証ガイドライン（2020.8 版）、プラント保安分野 AI 信頼性評価ガイドライン（2020.11 版）を参考にしている。

【保証（合意）の可否　凡例】
〇: 保証することができる（ただし保証を義務付けるものではない。）
△: 条件次第で保証することができる場合と保証することができない場合がある
×：保証することができない

	カテゴリ	内部品質特性	項目	保証（合意）の可否	備考
1	データ設計	要求分析の十分性	主要な品質低下リスクが発生する原因について検討を行い、記録すること	〇	ただし、検討結果の妥当性は保証できない。
2	データ設計	要求分析の十分性	検討結果に基づき、データの設計を行い、必要な属性等に反映すること	〇	ただし、検討結果の妥当性は保証できない。
3	データ設計	データ設計の十分性	システムが対応すべき様々な状況（ケース）を、網羅的に抽出すること（ex.自動運転のケース：「夜間の雨」「昼間の黄色」等のケースのうちシステムが対応すべきケースの組み合わせを抽出すること、ケースを絞り込むこと）	△	システムが対応すべきケース（状況）を網羅することができる場合は保証可。 ケースの組み合わせが多すぎる場合等、システムが対応すべきケースを網羅的に抽出することが難しい場合もある。この場合、網羅性を得ることを目指すことになる。

4	データ設計	データ設計の十分性	工学的な検討に基づき、属性の網羅基準（pair-wise coverage等 ）を設定し、その網羅基準を満たす属性値の組み合わせの集合をケースとして設定すること	○	
5	データの品質	データセットの被覆性／均一性	ケースごとに、元データから偏りのないサンプルを抽出すること（データセットに偏りがないこと）	×	収集されるデータに依存するため、偏りのないサンプルであることを保証することまではできない。ただし、可能な限り偏りが生じないよう抽出したことを示すことはできる。
6	データの品質	データセットの被覆性／均一性	偏りを生じさせないために行った活動を記録すること	○	
7	データの品質	データセットの被覆性／均一性	学習に必要なデータ量を確保すること	△	・第一次的にはユーザがデータを収集するため、基本的には保証不可。 ・ベンダがケースに応じたデータを収集することができる場合には、当該ケースを特定し、収集するデータ量の上限を決めれば保証可。ただし、「一定の性能を保証するためのデータを確保する」といった類の保証の仕方は不可。
8	データの品質	データセットの被覆性／均一性	合意した仕様に従い、アノテーションを施すこと	○	
9	データの品質	データセットの被覆性／均一性	複数のアノテーターが同一のデータについて別々にアノテーションを施すこと	○	
10	学習済みモデルの品質	モデルの正確性	合意した方法に従い、適切に前処理を施すこと、ラベル・正解値を付加すること	○	
11	学習済みモデルの品質	モデルの正確性	交差検証や汎化性能を確かめられる程度のデータ量を確保すること	△	ベンダがデータを生成、獲得することができることが前提。また汎化性能については、その目標値がある場合に保証可。
12	学習済みモデルの品質	モデルの正確性	訓練用・テスト用データの外れ値、欠損値の除去方法について検討し、実施記録すること	○	
13	学習済みモデルの品質	モデルの正確性	テスト用データを選定し、テスト用データの構成（検品の例でいえば、良品50%、不良品50%とする等）について設計すること	○	
14	学習済みモデルの品質	モデルの正確性	交差検証や汎化性能を確かめるための訓練データとテストデータを独立分離・管理すること	○	
15	学習済みモデルの品質	モデルの正確性	交差検証等、性能などの測定方法を確定させ、実施すること	○	
16	学習済みモデルの品質	モデルの正確性	適合率、再現率、F値等の評価指標を設定すること	○	通常「評価指標を達成すること」までは保証できないため、基本的には、評価指標を「設定すること」を保証しうるにとどまる。ただし、設定された評価指標値によっては、評価指標値を「達成すること」まで保証できる場合には、保証内容に含めることも考えられる。

17	学習済みモデルの品質	モデルの正確性	汎化性能を確保すること	×	汎化性能を確保するよう努力するものの、保証することまではできない。
18	学習済みモデルの品質	モデルの正確性	どのような汎化性能の測定が適切かを調査すること	○	
19	学習済みモデルの品質	モデルの正確性	汎化性能の目標値を定めること	○	ただし、「汎化性能の目標値を達成すること」は保証不可。
20	学習済みモデルの品質	モデルの正確性	汎化性能の測定方法を決めること	○	
21	学習済みモデルの品質	モデルの正確性	AUROCといった指標が十分であること	×	十分性を保証することはできない。
22	学習済みモデルの品質	モデルの正確性	学習後の正答率や損失関数の残差を収束させること	○	
23	学習済みモデルの品質	モデルの正確性	学習が局所最適に陥っていないこと	×	局所最適にならないよう努力するものの、保証することはできない。
24	学習済みモデルの品質	モデルの正確性	アルゴリズム選択やハイパーパラメータ設定を適切に行うこと	×	適切に行うよう努力するものの、適切性を保証することはできない。
25	学習済みモデルの品質	モデルの正確性	選択したアルゴリズムの選択根拠、ハイパーパラメータの設定根拠を明確にすること	○	
26	学習済みモデルの品質	モデルの正確性	どのようなハイパーパラメータを設定して検証したかを記録すること	○	
27	学習済みモデルの品質	モデルの正確性	公平性が要求される場合、公平性の比較手段を設定すること、対照テストの合格基準を定めること	○	
28	学習済みモデルの品質	モデルの安定性	ノイズ候補によりモデルの性能が著しく劣化しないこと	×	
29	学習済みモデルの品質	モデルの安定性	モデルに影響を与えるノイズ候補の洗い出しを行うこと。具体的には、誤差因子の選定とそれの与える影響解析を行うこと	○	ただし、抽出結果の妥当性については保証できない。
30	学習済みモデルの品質	モデルの安定性	敵対的データによる攻撃を防御する技術等近傍データ（元データに微小変化を加えて生成されるデータ）に対する安定性を評価する技術を適用すること	○	
31	学習済みモデルの品質	モデルの安定性	近傍データに対する安定性を保証すること	×	

32	学習済みモデルの品質	モデルの適切性	外部ライブラリのサプライヤ等の第三者との責任の所在を明確にすること	○	
33	実装・運用の品質	プログラムの健全性	信頼でき実績をもつソフトウェアを選定し、その選定経緯を記録すること	○	
34	実装・運用の品質	プログラムの健全性	選定したソフトウェアについて、その欠陥の発見等を運用期間中にモニタリングし、必要に応じて修正等の措置をとること	○	
35	実装・運用の品質	プログラムの健全性	シミュレータを活用するなど、ソフトウェアの健全性の維持に必要な保守体制を構築すること	○	
36	実装・運用の品質	プログラムの健全性	バリデーションおよびテストフェーズにおいては、原則として実用段階で用いられる計算環境を模倣した環境でバリデーション・テストを行うこと	○	
37	実装・運用の品質	運用時品質	外れ値に対する学習除外の仕組みを実現すること（オンライン学習の影響を考慮すること）	○	
38	実装・運用の品質	運用時品質	外れ値を監視すること（モデルを更新するデータが想定したデータ区間を外れているかを監視するなど、入力データの質を監視すること）	○	
39	実装・運用の品質	運用時品質	モデルの品質劣化・誤判断についてモニタリングすること	○	
40	実装・運用の品質	運用時品質	デグレードを許容可能な範囲に収めること	×	許容範囲内に収めるよう努力するものの、保証することはできない。
41	実装・運用の品質	運用時品質	実運用でしか収集できないデータを記録する仕組みを構築すること	○	
42	実装・運用の品質	運用時品質	オンライン学習を行う場合には、予想外の品質低下がもたらす影響について検討し、必要な場合には動作範囲の限定などのシステム上の対応をとること	○	
43	実装・運用の品質	運用時品質	経年劣化の進行が早いモデルであると想定される場合、これに合わせてモデルのチューニング頻度を設計し、これに従い実行すること	○	ただし、設計した頻度の妥当性は保証できない。
44	実装・運用の品質	運用時品質	実データに対する予測品質が劣化しないこと（モデルを陳腐化させないこと）	×	
45	実装・運用の品質	運用時品質	訓練データの特性変化や出力の追加等により再学習を行った結果、再学習前の性能に対する劣化をあらかじめ定めた許容範囲内に収めること	×	
46	実装・運用の品質	運用時品質	モデルの更新を自動で行う場合に、モデルの特性変化や性能変化が許容範囲であることを検査できる仕組みを講じること	○	
47	実装・運用の品質	運用時品質	学習後の正答率や損失関数の残差を収束させること	○	
48	実装・運用の品質	運用時品質	どのようなハイパーパラメータを設定して再学習をおこなったかを記録すること	○	
49	実装・運用の品質	運用時品質	学習用データセットのバリエーションが増えた場合も検証ができるように、交差検証の方法を定めること。交差検証を実施すること	○	
50	実装・運用の品質	運用時品質	再学習時や追加学習時に、交差検証や汎化性能を確かめるための訓練データとテストデータを独立分離・管理すること	○	

5. プロダクト種類毎の対応例

プロダクト事例別に、特に問題となる品質特性と具体的な対応の例および留意点を示す。

5-1 不良品検査

外観をカメラで捉え、画像（動画）データを処理し、不良品・良品、異常の有無の判定、または不良の個所の特定等を行う。食品原料や工業製品の不良品検査

（1） 特に問題となる品質特性

「性能（有用性）」

例 1： 目標値として、パフォーマンスを測る指標を決定する。「精度」という曖昧な表現での合意はなく、指標として「正確性」「適合率」「再現率」のうち、何を評価するのかを定義、決定し、その指標について「目標値」の認識を合わせる。

例 2：指標の基準値については、AI 導入の目的（検査業務の一部置き換え（検査人数が減る等）なのか、作業員の支援（可視化）なのか等）について改めて認識を合わせ、現状の作業人員による精度と照らし合わせて現実的な目標値を設定する。

例 3：流れ作業を対象とする場合等、推論を常時行う必要性やライン速度の制約があるため、精度に加えて推論の速度についての確認も重要となる。

「頑健性」

例:装置の経年劣化による精度劣化の可能性を考慮して、運用で対応するべきことを明確にして示す。（例：ベルトコンベアに汚れが付着しないよう清掃すること）

（2） その他特に留意するべき事項

- 作業員の現状の作業に照らして、AI 導入のメリットをどのように出せるかについてよく議論する。自動化の精度を挙げるという議論の前に、どのような部分で効率化を測るかといった利用方法についてよく話し合う。
- 訓練データ（多くの場合は新しく収集する必要）について、現場の状況により近い状態のデータどのように集めるかについて認識を合わせる（必要に応じて現場ヒアリング等を実施）。また運用時データとの整合性をどの程度取れるか、について認識を合わせる。
- 実際にシステムを使う現場（工場等）とのコミュニケーションを密にし、現場との協力体制を取りつつ進める。
- AI 単体で利用されるのではなく、検査システム全体の一部である場合は、全体システムとの関連性で議論するべき項目の明確化・調整が必要になる。

5-2 ひび割れ検出

建物や公共物の損傷個所を点検する外観を画像に撮影し、その画像上でひび割れ個所を検出する。

（1） 特に問題となる品質特性

「性能」

例：損傷箇所の検知の精度を測る指標には、各社独自の基準の他に、コンソーシアム等により共通の評価基準を検討する動きがある。分析の結果、損傷箇所の見落としを極力少なくすることが要求されるため、「解像度○○mm/ ピクセル以上の画像において、幅○○mm 以上のひび割れを○% 検出する」という再現率で性能を示す方法がある。

「頑健性」

例：示した精度で検出するためには、画像の撮り方が重要となる。解像度、気象条件、画像の撮り方について、前提となる条件の確認をする。撮り方と解析方法は、セットで考える必要がある。

（2） その他特に留意するべき事項

- インフラの点検用途における AI の活用は、人による点検の効率化のためのスクリーニングが目的となる。詳細な点検は足場をかけて人間の目視や打音検査が必要であり、その詳細点検をする前の業務の効率化が目的であるということへの認識を合わせる。
- 何を損傷として検出するかの認識が合わないと、成果物の納品時点でのトラブルになりやすい。検出する損傷の定義（色や形状等）、どのような資料を納品するかについて、前もって認識を合わせることが重要。

5-3 株価予測

過去の株価の変動実績を下に、関連する要素を処理し、将来の株価の変化を予測する。

（1） 特に問題となる品質特性

「性能（有用性）」

例 1：開発契約時に PoC と同様の性能がでることの保証は難しい旨を確認し、保守運用フェーズにおける監視方法など、どのようにモニターしていくかについてを取り決める。例えば、保守運用の中で監視すること、指標で設定した数値を下回った際に議論をする等、保守契約の条項のなかでプロセスについて合意する。

例 2：運用時にモニタリングする項目としては、株価が「上がる」としたもののうち、何 % が本当に挙がっているか、というような数値があり得る。PoC の時に学習したバックテストの数字を下に決する。

例 3：PoC で検証していることを踏まえて、ある一定の前提条件をつけた上での目標値を努力義務として示す。または、運用時に期待する結果が得られないケース等において、想定される原因と改善策の調査・説明を努力義務として対応する。

(2)　その他特に留意するべき事項

- 過去データをどの期間までカバーするかについてについての認識を合わせることが重要。過去の株価市場の環境をみて、大きなイベントをどの範囲でカバーするかによって 5 年前なのか、10 年前なのかの期間が決まる。期間が長ければ良いというわけではなく、加味したいイベントが含まれているかの視点が重要であり、専門知識が必要になる。

5-4 EC サイト

EC サイト上で、閲覧ユーザに対して購買商品のレコメンドを行う

（1）　特に問題となる品質特性

「性能」

例 1：マーケティングの文脈で、導入後のオンライン効果の検証が重要となる。ウェブサイトの AB テストと同様の比較や購入数の向上等について目標値の設定により、性能の評価をしていくことが有効。ユーザグループごとに、もともとの購入数のばらつきがないようにするなど、適切な効果を測るための工夫も必要になる。

例 2：新商品の割合などによっても予測値（購買率等）との差分は変化する等、実際の効果の確認には、売上げの推移等を時系列で一定の期間見ていくことが必要になる。契約の期間を区切り、期間毎に暫定的な KPI を設定し前の期間と比較しつつ、徐々に向上を目指すプロセスを踏む。

例 3：既存データをベースにしたオフラインの評価方法には、レコメンド独自の評価方法（Precision@K、MAP 等）を使って一定の効果を測定する方法もある。ただしこれらのオフライン評価では目的（売り上げ向上等の）を達成しにくい場合も多く、実務上はオンライン評価で効果をみていくことが有効。

（2）その他特に留意するべき事項

- 購買商品のレコメンド機能には、過去データの分析による購買予測や同様のものばかりをレコメンドするといった機能ではなく、潜在的な欲求を拾ったり、ある程度のランダム性を狙ったりといったニーズがある。導入後にもマーケティングの観点で分析を重ねながら業務に落としていく視点が必要になる。

※他のレコメンド系 AI に係る留意点：「公平性」
人を対象にして判定を行うような機能を持つ場合（例えば、採用エントリーシートの審査支援や、保険商品のレコメンド等）は、「公平性」も重要な留意点になりうる。
「公平性」への対応例としては、判断材料として含みたくない属性や要素があるかを検討し、除きたい属性がある場合はそれを特定し、データ属性から外す等が考えられる。
公平性をどのように実現したいかは、AI を使う領域における全体の方針や戦略、管理項目にも大きく関わる。ユーザ側での方針の整理と決定が重要となる。

5-5 医療画像診断

レントゲン検査や CT 検査で取得した画像を読み取り病気や骨折といった症状の判別を行う。

（1） 特に問題となる品質特性

「性能」

例 1：再現率と適合率どちらの指標を優先するか運用面も考慮して決定する必要がある。医療画像診断の場合は一般的には病気や骨折といった症状の見落としが問題となることも多いが、再現率の指標を優先する（つまり見落としは絶対に防ぐ）事に集中し、誤検知の件数が増えてしまっても、病院業務の支障をきたす可能性が出てくる。そのため、運用を最優先に再現率と適合率のバランスを図ることが必要。

例 2：どのような医療機関、どのような医療シーンで使われるかを意識することも大事である。具体的には健診センターのような、必ずしも病気や骨折といった症状を把握されていない方の健診の際に使われる可能性もあるし、病院で使われる場合でも一次医療から三次医療、それぞれ医療体制が異なるケースで使われる可能性がある。その結果、見落としを防止することが重要なのか、誤検知を減らすことが重要なのか、それぞれの病院の立ち位置で要件が異なる可能性があるので、AI の目的を事前に整理することが重要。

「頑健性」

例 1：正常、異常それぞれの学習データをより多く集めていくことが AI の品質に寄与する可能性が高い。しかしながら学習データが集められれば良いわけではなく、例えば診断画像の部位に人工物が挿入されているケースなど、患者様独自の症状によって、頑健性が保てない可能性も否定できない。品質を定義する前にそのような状況を品質指標と対象外として扱うのか、対象として扱うかを協議する。

例 2：医療画像を作成する医療機器の変更、または医療機器の設定によって精度が落ちる可能性がある。一般的に医療画像は DICOM と呼ばれる業界フォーマットとして出力される為、見た目としては医療機器による差分は少ないように見えるが、実際の画像としてはコントラストに差が出るケースも多い。そのため、特定の医療機器、及び設定で取得したデータを教師データとしていた場合、その他の医療機器を使う事で頑健性が保てないケースもある為、対象医療機器の定義などを実施する

「セキュリティ」

例：医療画像の中に患者情報が残っている可能性を考慮する必要がある。一般的には医療画像の提供を受ける場合は患者情報などの個人情報が含まれないという前提条件を設けることが多いが、実際に提供された医療画像の中には、患者情報が含まれてしまうこともあり得る。このような意図しない形で患者情報が含まれていた時、またその患者情報に気づかない状況で開発が進んでしまった場合のリスクを事前に協議する、もしくは契約書上でリスクを回避する。

（2） その他特に留意するべき事項

- 作成された AI が実際にどのような医療機関で、どのような運用下の中で使われるかを定義した

上で、性能、頑健性、セキュリティを意識する必要がある。また要件を決めていく中では、AIで全ての要件を解決するのではなく、仮に AI が誤った判断を出力した場合でも医療業務に影響を与えないような運用を検討することが望ましい。

5-6 自動応答

ユーザからの質問をもとに、FAQ などの事例情報の類似検索を行い回答するチャットボットや、音声認識や音声合成と組合せてコールセンター業務を行うボイスボット等。業務としては、本人確認を行い、必要な情報を応答する。

（1） 特に問題となる品質特性

「性能（有用性）」

例 1：実利的な FAQ を行うチャットボットでは、チャットボット上でのユーザからの満足度回答などもとに導入効果を算出する。ボイスボットでは、ボットによる回答率、（切電や人間のオペレータを呼び出した離脱率）などを算出する。いずれの場合も、実際のユーザとのインタラクションで答えが大きく変わるため、PoC 段階での数値はほとんどあてにならない。

例 2：特にボイスボットでは、音声認識の結果もバラツキが大きく、一定の上限、下限を設定することは難しい。数値目標の設定が難しいことから、呼量を絞ったうえで、実際のエンドユーザにサービスを提供するパイロット運用期間を設定する。この間にモデルを改善しながら、本番サービスへの導入可否を判断してもらう契約とする。

「性能（リスク回避性）」「公平性」「セキュリティ」

例：チャットボットにキャラクター性を持たせる場合などでは、Web データで大規模に事前学習した生成系の DNN を使うことも考えられるが、こちらは差別的な表現をしないかなど、リスク評価について顧客と認識を合わせることが極めて重要

（2） その他特に留意するべき事項

- ボイスボットなどのような、多くの要素が絡むシステム、ユーザとのインタラクションがキーとなるシステムの場合、PoC 段階では実環境の多様性（様々な話者、会話のスクリプトによる反応の違い）を十分に評価することはほとんど不可能である。安易に数字を定めず、パイロット運用期間での改善手段を多く残すよう、設計段階から検討する。
- 無人化ではなく、省力化を目指すこと、構築、検証、運用を進める中での課題解消や運用時の再学習などを通じて、省力化の度合いを高めていくような交渉を行う。

5-7 無人搬送車（AGV）

倉庫や工場内で、荷物を自動で搬送する。ここでは無軌道の場合を想定する。

（1） 特に問題となる品質特性

「性能（有用性）」

例 1：倉庫や工場内の搬送作業は複数のタスクが連携している（例えば、高いところから荷物を降ろして床に置く作業、A 地点から B 地点に運ぶ作業等）。AGV 導入には、プロセスの全てを対象にするのではなく、タスクを区切って整理して、自動化で効率を図りたいタスクを明確にする。

例 2：搬送プロセスにおける物量と AGV のスペック（スピードや可搬重量等）により、AGV の必要台数を設定するが、無人化による効果は、導入台数＝削減人数ではなく、自動化をしたタスクにおける作業時間の削減をみる。

「頑健性」

例：生産ラインや倉庫レイアウトの編成や、季節による積み荷の変化などのパターンを十分に考慮してシミュレーションによる確認を行う。

（2） その他特に留意するべき事項

- 倉庫・工場内の搬送作業はタスク毎に複数の機能が合わさって全体のプロセスを構成している。それぞれのタスク毎に自動化が適するか否かを性能（有用性）、頑健性、安全性等の視点を加味した上で各タスクの効果を整理し、自動化の範囲をどのように区切ってすみ分けるかが重要となる。

「契約締結における AI 品質保証の在り方」研究会

■　話題提供者リスト（研究会開催順）：

石川冬樹氏（国立情報学研究所）、丸山宏氏（株式会社 Preferred Networks）、黒河徹也氏（株式会社調和技研）、小峰弘雅氏（株式会社ベイカレント・コンサルティング）、大岩寛氏（産業技術総合研究所）、古川直裕氏（株式会社 ABEJA)、後藤大氏（晴海パートナーズ法律事務所）、駒村和彦氏（株式会社野村総合研究所）、太田満久氏（株式会社ブレインパッド）

■　ヒアリング協力先リスト

株式会社 AVILEN、株式会社イクシス、エッジテクノロジー株式会社、AnyTech 株式会社、株式会社調和技研、HEROZ 株式会社、富士ソフト株式会社、Musashi AI 株式会社、
藤吉弘亘氏（中部大学）、山下隆義氏（中部大学）

「契約締結における AI 品質保証の在り方」研究会メンバーリスト

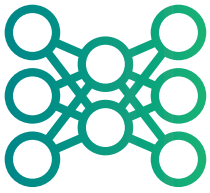
座長：南野充則（株式会社 FiNC Technologies/JDLA 理事）

副座長：柿沼太一（STORIA 法律事務所）、後藤大（晴海パートナーズ法律事務所）、福岡真之介（西村あさひ法律事務所）、古川直裕（株式会社 ABEJA）、渡邊道生穂（HEROZ 株式会社）

研究員（五十音順・敬称略）：

秋元一泰（華為技術日本株式会社）、石川冬樹（国立情報学研究所）、大利優（野村ホールディングス株式会社）、落合孝文（渥美坂井法律事務所・外国法共同事業）、加藤奈穂（株式会社調和技研）、金正福（株式会社調和技研）、工藤郁子（世界経済フォーラム第四次産業革命日本センター）、黒河徹也（株式会社調和技研）、小峰弘雅（株式会社ベイカレント・コンサルティング）、小宮山正樹（エッジテクノロジー株式会社）、砂金優介（ジャパニアス株式会社）、但野友美（株式会社調和技研）、林憲一（華為技術日本株式会社）、松本清一（有限責任監査法人トーマツ）、八木聡之（富士ソフト株式会社）

本ハンドブックで記載する内容は、研究会における議論によるものであり、特定の企業や組織の意見を代表するものではありません。



Japan
Deep Learning
Association

https://www.jdla.org/