



東日本大震災から14年 福島事故以降の安全学について

東北大学大学院 工学研究科技術社会システム専攻 教授 たかはし まこと 高橋 信氏

2011年の東京電力福島第一原子力発電所の事故以降、原子力安全規制体制の見直しや新安全規制基準の導入など、原子力発電に対する安全規制の強化が図られ、事業者においても自主的な安全性向上の取り組みを進めてきました。

安全とは何か。著者は安全に関わる認識と取り組みについて、事故以前と根本的に異なる視点を持つ必要がある、と指摘します。社会技術システムの専門家として、また政府・自治体の原子力委員などを務める立場から、著者はあらためて「安全とは何か」について深く考察します。

東京電力福島第一原子力発電所の事故（以下、福島事故）から14年が経過しました。この間、東日本におけるBWR（沸騰水型軽水炉）原子力プラントの再稼働は停滞していましたが、最近になり東北電力が先陣を切つてようやく進展し始めています。しかし、原子力の安全性に対する懸念は依然として根強く、新たなプラント建設に踏み切るまでには至っていません。

福島事故以降、各電力事業者は原子力規制委員会が定めた「新安全規制基準」に基づき、原子力プラントの安全性向上に取り組んできました。莫大な資金を投じて新しい安全設備を導入し、既存施設を更新するだけでなく、安全文化の向上にも力を注いでいます。その結果、福島事故以前と比べてプラントの安全性は大幅に向上しているといえます。それでもなお、社会における原子力への不信感や風当たりは厳しいままです。

福島事故による住民避難がまだ続いている現状を目の当たりにすれば、国民の不安や不信感は理解できるものです。原子力発電所が「安全」であると強調してきたにもかかわらず、住民の生活を根底から揺るがす事故を起こしてしまったという事実是否定できません。原子力発電を今後も国の基幹電源とするためには、安全への取り組みを福島事故以前とは根本的に異なる視点で進める必要があります。そして、その過程で国民の理解を得る努力を怠ってはなりません。

本稿では、安全に対する新しい視点である「Safety-II」の

概念を中心に、人間を基軸とした安全性向上の取り組みについて紹介します。

安全への取り組みの変化

歴史的に見ても、安全性を高める努力は常に続けられてきました。その結果、100年前と比べて、私たちの生活は格段に「安全」になっているはずですが、もちろん、東日本大震災のような大規模な災害は例外ですが、日常生活における事故や災害のリスクは大幅に減少しています。

しかし、多くの人が「以前より安全になった」と実感できていないのも事実です。この理由の一つは、災害や事故の頻度が減ったことで、一度発生するとその出来事が強く印象に残るからです。特に、未曾有の災害を経験すると「安全になった」という主張は説得力を持ちにくくなります。

それでも、従来の安全対策により私たちの社会は確実に安全になってきました。ただし、その一方で社会システムがますます巨大化・複雑化したため、これまでとは異なる視点で安全を考える必要が出てきています。それが「Safety-II」という考え方です。この概念を説明する前に、私たちが対象としている「社会技術システム(Socio-Technical System: STS)」について説明します。

社会技術システムとは、高度な技術システム(機械や装置)と、それを管理・運用するチームや組織(社会システム)

【図1】



社会技術システムの例

航空機の運行管理

航空管制

製造工場

ネットワーク管理

鉄道の運行管理

手術室でのオペレーション

出典：著者プランにより作成

Safety-IIでは「何がうまくいっているのか」に着目し、それを広く共有することで安全性を向上させようとしています。例えば、ある組織で事故率が⁴10(1万回に1回)だった場合、Safety-Iではその1回の事故を徹底的に分析します。しかし、Safety-IIでは、残りの9,999回の「成功事例」にこそ学ぶべきポイントがあると考えます。

この考え方を理解するには「手順通りにやればうまくいくのが当たり前」という発想を変える必要があります。現場で業務が適切に遂行されているのは、単にルールを守っているからではなく、現場の人が適切に調整しているからです。この調整こそが、安全を支えている重要な要素であり、Safety-IIの基本的な考え方になります。

Safety-Iによって安全性はある程度向上しました。しかし、さらにその先を目指す組織では、Safety-IIの視点が重要になります。これは単なるコンプライアンスの強化ではなく、新たな形で安全性を高める可能性を持っています。現場の良好な取り組みに注目し、それを組織全体で共有・活用することによって、より実効性の高い安全文化を築くことができるのです。

安全性の指標(Leading Indicator)の重要性

ここでは、「安全」をどのように評価するかについて考えます。

ム)が密接に結びついたものです。例えば、原子力発電所、鉄道や航空といった輸送システム、電力の発送電システムなど、現代社会を支えるインフラの多くが該当します【図1】。

こうしたシステムでは、単なる機械の故障や人間のミスだけでなく、さまざまな要因が複雑に絡み合っつて事故やトラブルが発生します。そのため、従来の単純な管理手法では対応が難しくなっています。システムが複雑であっても、その機能が個別に分解できるなら「扱いやすい(tractable)」システムです。この場合、問題を細かく分けて解決する従来のアプローチ(要素還元主義)が有効でした。しかし、現代の社会技術システムは、要素還元主義が通用しにくい「扱いにくい(intractable)」システムになっています。これが、従来の安全管理(Safety-I)だけでは十分でなくなった理由の一つです。

次に従来の安全管理(Safety-I)の限界について考えてみます。Safety-Iとは「トラブルが発生した原因を特定し、それを取り除くことで再発を防ぐ」という考え方です。このアプローチは、1970年代までの「扱いやすい」システムでは非常に効果的でした。例えば、トラブルが起きた際に「原因を見つけて修正する(Find-and-Fix)」ことで、安全性は大きく向上しました。現代の社会技術システムでは、安全性をさらに高めるために、手順を増やしたり、ルールやペナルティを厳しくする方法が取られています。しかし、これらの施策の効果には疑問が投げかけられています。では、Safety-IIとはどのような考え方なのでしょうか？

前節で紹介したSafety-IIの考え方では、成功が多い状態を「安全」と捉えます。しかし、現在の一般的な安全評価基準では、失敗やトラブルが少ない状態が「安全」と見なされることがほとんどです。実際、ISOにおける安全の定義は以下のとおりです。

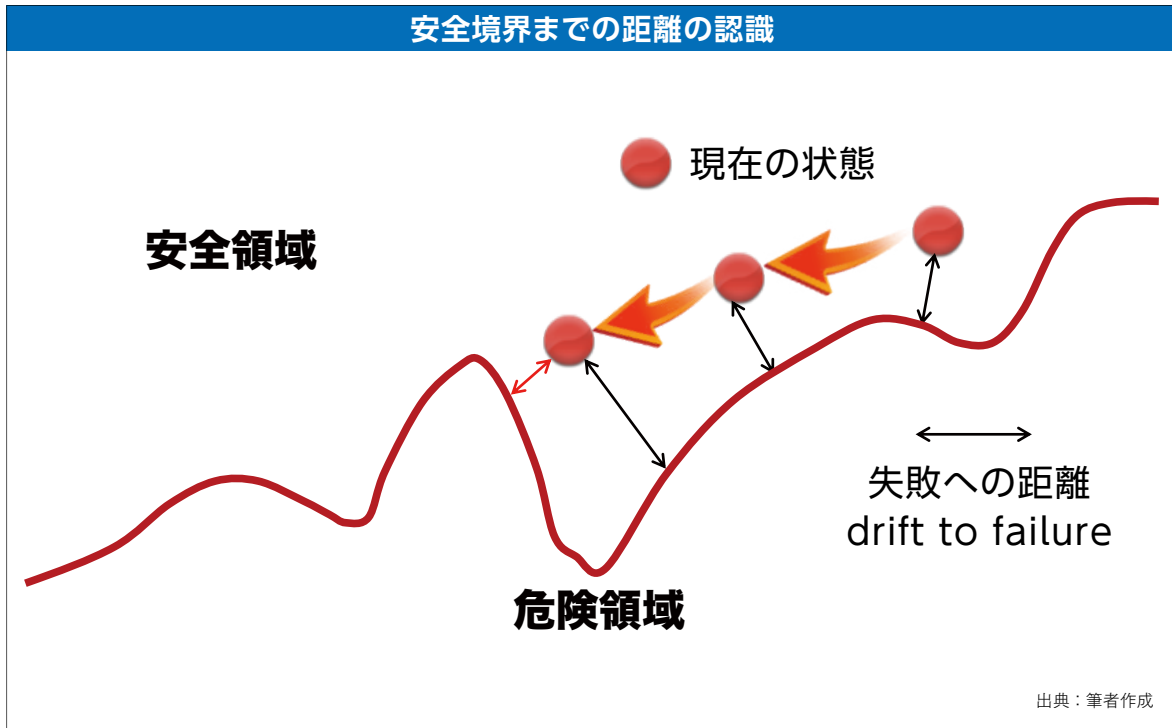
Safety is a condition where the number of adverse outcomes (accidents / incidents / near misses) is as low as possible.

安全とは、好ましくない結果(事故・インシデント・ニアミス)の数が可能な限り少ない状態である。

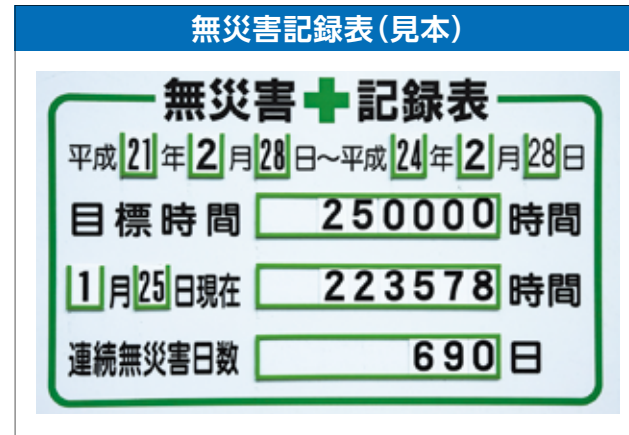
この定義に基づけば、安全を高めることは、事故・インシデント・ニアミスの数を減らすことと同義になります。この定義に異議はありません。実際、不安全な事象が少ないことは、安全であることを意味します。しかしながら、不安全な事象の数だけに注目し、それを減らすことばかりが強調される現状には、見直すべき点があるのではないのでしょうか。

例えば、多くの産業の現場では「1年間無事故が続けば、その部門は安全であり、優秀である」と判断されることが一般的です。そして、無事故記録を維持するように厳しく指導されることも少なくありません【図2】(8ページ)。しかし、1年間無事故が続いたという事実と、その組織が現在も安全であるということは、必ずしも同じではありません。確かに、1年間事故がなかったということは、安全意識が一定の

【図3】



【図2】



レベルにあることを示唆しますが、「たまたま」事故が発生しなかった可能性も考えなければなりません。

実際、2010年にメキシコ湾原油流出事故を引き起こしたBP社の海上石油掘削施設「Deepwater Horizon」は、事故発生までの7年間、無事故記録を保持していました。このような過去の統計に基づく安全性評価は「遅れ指標(Lagging Indicator)」と呼ばれ、過去の事実を示しているにすぎません。しかし、安全評価において本当に重要なのは、現在、そして今後の安全性です。そのためには「先行指標(Leading Indicator)」が必要になります。この先行指標をどのように評価するかは、難しい問題であり今後の重要な研究テーマといえるでしょう。

Safety-IIでは、組織の状態は常に変化し、長期的な流れの中で「失敗への漂流(drift to failure)」というモデルをもとに考えます。これは「組織の状態は常に動いている」という前提に基づいています。例えば、安全対策に十分な資金や人材を投

入し、その結果、安全性が向上したとします。そのとき多くの組織は、その状態が維持され、安全性の高い状態が続くと考えがちです。しかし、現実には、安全な状態がそのまま維持されることはきわめてまれです。さまざまな要因により、組織の状態は常に変化し、往々にしてその変化は「安全性の低下」という方向に向かいます。この変化を「漂流(drift)」と表現し、安全領域から外れたときにトラブルが発生すると考えます【図3】。

この「drift to failure」モデルが示唆する重要な点は、大きく分けて2つあります。

① 安全境界までの距離の認識

安全を管理する上で重要なのは、組織の現在の状態と「安全境界までの距離」を適切に認識することです。具体的には、

- ・現在の位置が安全境界に対してどこにあるのか。
- ・どの方向へ進んでいるのか(安全の淵に向かっていくのか、それとも並行に動いているのか)。
- ・変化の速度はどの程度か。

これらの要素を把握することが、安全管理においてきわめて重要です。例えば、東京電力福島第一原子力発電所のケースでは、一見すると適切な安全管理が行われているように見えました。しかし、実際には、大規模地震による津波や、それによる外部電源喪失といった事象に対して脆弱であり、

結果的に「失敗境界までの距離」が短かったと考えられます。

② 失敗境界までの距離の認識の難しさ

もう一つの重要なポイントは、組織が「失敗境界までの距離」を正確に認識することが難しい、という点です。これは、以下の理由によりります。

- ・漂流している船からは、どこが「point of no return (引き返せない地点)」なのかを正確に把握することが難しい。
- ・人間の持つ認知バイアスが影響し、安全リスクの判断がゆがめられる。

実際、組織のリーダーや管理者が「私たちの職場は安全だ」と確信している場合には、認知バイアスによって実際のリスクを過小評価している場合が多くあります。こうしたバイアスを完全になくすことは難しいですが、少なくとも「私たちは本当に安全なのか？」と常に問い直し、リスクに対して注意を向け続ける姿勢が重要です。

リスクを認識し、安全性を向上させる主体である人間の特性

ここまでは、安全に対する新しい視点について述べてきましたが、本節では、安全性を評価する側である「人間の

認知バイアスについて述べます。

安全性を高めるための基本となるのはリスク評価です。私たちは、対象とするシステムのリスクを評価し、対応策を検討します。しかし、このリスク評価は多分に主観的にならざるを得ません。なぜなら、安全性向上の活動の基本となるのは人間が感じる「主観的リスク」だからです。言い換えれば「危なそうところ」を見つけることが安全対策の第一歩ですが、その「危なそう」という感覚は、あくまで主観的なものです。もちろん、事故やトラブルの発生頻度などの客観的なデータに基づいてリスクを評価することは重要です。しかし、前述のとおり、これらの間接的な指標とリスクの関係は必ずしも明確ではありません。そのため、人間の主観的なリスク認識には、さまざまなバイアスが影響を与えることになります。「危なそうところを見つける」と聞いて、現場で行われるTool Box Meeting（安全ミーティング）などの活動を思い浮かべる方もいるかもしれません。しかし、ここで述べる主観的リスク評価における認知バイアスは、こうした活動で重要視される「危険感受性」とは異なるものです。

リスク認識に影響を与える代表的な認知バイアスには、以下の3つがあります。

◎ 確証バイアス

自分にとって都合の良い情報ばかりを無意識に集め、反証となる情報を見たり、集めようとしなかったりする傾向。

が起こるのではないかと考えてしまうのです。

これらの認知バイアスは、すべての人が持つ傾向であり、必ずしもネガティブな側面ばかりではありません。進化的に考えれば、人類がこれまで生存してこられたのは、これらのバイアスが生存戦略として有効に機能してきたからともいえます。したがって、認知バイアスを完全になくすることはできません。重要なのは、バイアスがあることを前提に考え、適切に対処することです。

原子力の安全を考える上で、特に重要なのは「近接性バイアス」ではないでしょうか。このバイアスは、リスク対策の優先順位に影響を与えます。

例えば、東日本大震災直前の原子力業界の状況を振り返ると、2007年の新潟県中越沖地震の影響で「耐震裕度」の問題にリソースが集中していたことを記憶されている方も多いと思います。しかし、当時「耐震」の問題と同じくらい、あるいはそれ以上に重要だった「津波」のリスクに十分な意識が向けられませんでした。これは「津波のリスクは低い」という信念による確証バイアスの影響が関係していた可能性がります。その結果、津波リスクを示唆する情報が軽視され、対策が間に合わなかったと考えられます。もちろん「事前に対策を取るべきだった」と単純に結論づけることはできません。そのように考える人もいますが、それは「後知恵バイアス」による影響を受けている可能性があります。

現在の原子力業界も、福島第一原子力発電所の事故によ

◎ 正常性バイアス

多少の異常が発生しても、それを「正常の範囲内」と判断し、深刻な問題として受け止めない傾向。

◎ 近接性バイアス

直近の出来事が記憶に強く残るため、同じような事象が高い頻度で再び発生すると感じる傾向。

リスク評価を行う際、人は無意識のうちに認知バイアスの影響を受けます。例えば、あるシステムが「安全である」と思い込んでいる場合「安全であることを示す情報」ばかりに注意が向き「リスクを示す情報」を無視する傾向があります。これが確証バイアスによる主観的リスク認知のゆがみです。確証バイアスは、特にインターネット上の情報を参照する際に顕著に現れるといわれています。現在、社会問題となっている「分断」は、この確証バイアスが情報の偏りを生み、それを助長することで発生しているとも考えられています。

リスク評価の偏りを生むもう一つの要因が正常性バイアスです。これは「少しくらいの異常は正常の範囲内」と捉えてしまう傾向を指します。リスクが徐々に増大している状況においても、人は「まだ問題ない」と判断しがちになるのはこのバイアスの影響と考えられます。近接性バイアスもリスク評価に大きな影響を与えます。このバイアスは「直近で起こった出来事が、再び高頻度で発生すると考える傾向」を意味します。例えば、大きな災害が発生した直後には、その災害に関する報道が頻繁に行われるため「また同じ災害

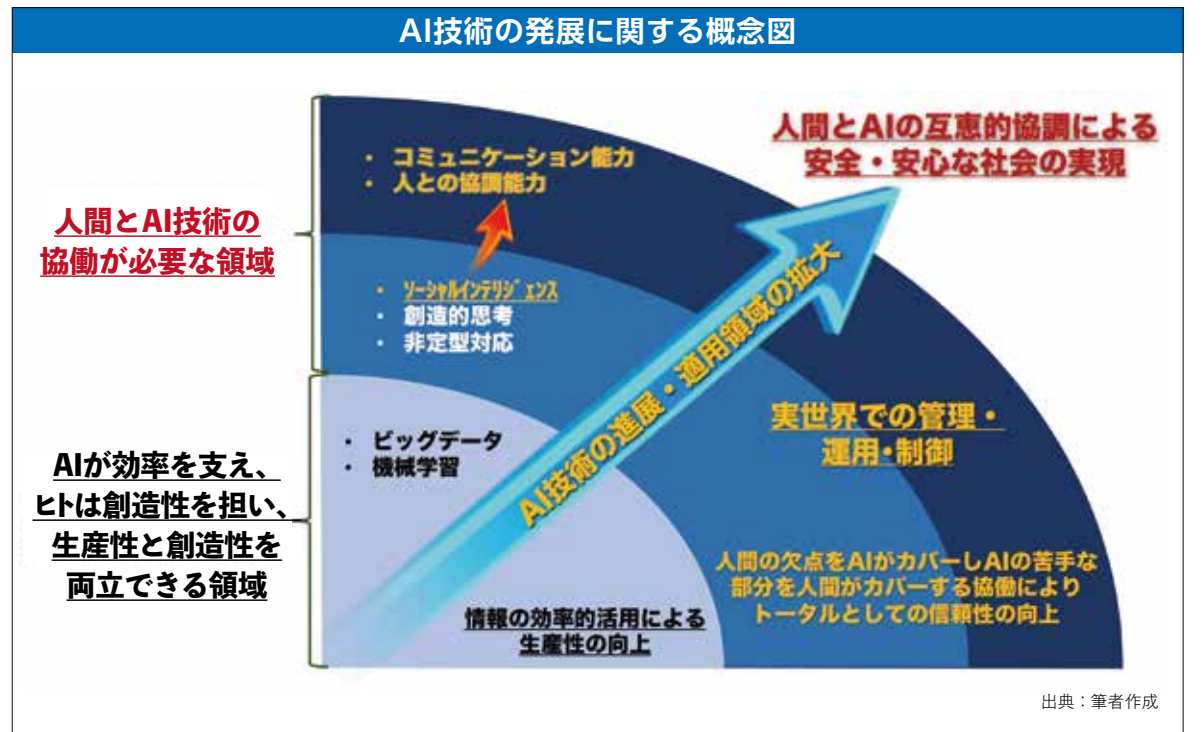
る「近接性バイアス」の影響を受けているのではないのでしょうか。震災から14年が経過し、新たな安全規制基準のもとで安全対策が実施され、発電所の安全性は大幅に向上しました。しかし、多くの安全対策は「福島第一原子力発電所の事故を繰り返さない」ことに重点が置かれています。これは、原子力を再び社会に受け入れてもらうための必要条件ではありますが、それだけで十分でしょうか。

運転訓練のシナリオを見ても、過酷事象として導入されているのは福島事故に類似したケースがほとんどです。しかし、次に起こる重大事故は、福島事故とはまったく異なる様相を呈するかもしれません。私たちは、そうした未知のリスクに対しても備えをしておく必要があります。それは、必ずしも大規模で高価な安全システムを導入することではなく、柔軟で効果的な対策を講じることが重要なのです。

AIへの過度の依存への警鐘

近年、将棋や囲碁においてAIが名人に勝利したことはAI技術の進展を象徴する出来事であり、人間を凌駕する技術として強い印象を与えました。しかし、こうしたゲーム分野での成功と、一般社会での問題解決におけるAIの優位性を同一視することは適切ではありません。囲碁や将棋はルールが明確で不確実性が少ない環境での意思決定を要するため、AIが最も得意とする分野です。

【図4】



データを漏れなく正確に扱うことが要求される安全管理の効率化には、AI技術が大いに貢献できる可能性があります。

おわりに

本稿では、安全に対する新しい視点である「Safety-II」の概念を中心に、人間を基軸とした安全性向上の取り組みについて紹介しました。従来の取り組みである「Safety-I」とは異なる視点で安全を捉える「Safety-II」の概念、安全を担う主体である人間の避けることのできない認知バイアスを意識することの重要性を最後に強調したいと思います。AI技術は現状では安全を担う人間を代替できるほどの信頼性を有してはいません。今後しばらくはAI技術の支援を受けながらも人間が社会技術システムの安全の要として位置づけられる状況が続くでしょう。

◎参考文献

- ・レジリエンスエンジニアリング―概念と指針／北村正晴(監訳) 日科技連 2012
- ・実践レジリエンスエンジニアリング／北村正晴・小松原明哲(監訳) 日科技連 2014
- ・レジリエンスエンジニアリング応用への指針―レジリエントな組織になるために／北村正晴監訳 日科技連 2017
- ・Safety-I & Safety-II―安全マネジメントの過去と未来／エリック・ホルナゲル(著)、北村正晴・小松原明哲(監訳) 海文堂 2015
- ・Safety-2の実践―レジリエンスポテンシャルを強化する／エリック・ホルナゲル(著)、北村正晴・小松原明哲(監訳) 海文堂 2019
- ・ザ・セカンド・マシン・エイジ／エリック・プリニョルフンほか(著) 日経BP 2015
- ・HUMAN+MACHINE 人間+マシン―AI時代の8つの融合スキル／ポール・R・ドーアティほか(著)、保科学世(監訳) 東洋経済新報社 2018
- ・そろそろ、人工知能の真実を話そう／ジャン・カブリエル・ガナシア(著) 早川書房 2017
- ・AIvs教科書が読めない子どもたち／新井紀子(著) 東洋経済新報社 2018
- ・人工知能 人類最悪にして最後の発明／ジエイムス・バット(著) タイヤモンド社 2015

一方、ゲーム以外の分野においてもAIは着実に進展しています。たとえば、過去のデータをもとにした予測(寿命予測、保全最適化、投資判断)、画像データの分析(生体認証、状況判断)、さらには文章やイラストの生成といった応用があります。これらの領域では実用的な成果が数多く生まれています。しかし、これらはあくまで限定された問題領域での応用であり、俯瞰的な判断が求められる社会技術システムの安全性向上にそのまま適用できるわけではありません。

【図4】は、AI技術の発展に関する概念図です。現状では、情報を効率的に活用することで生産性を向上させる方向でAI技術が活用されています。一部の成功事例では、AIが効率性を支え、人間が創造性を担うことで、生産性と創造性の両立が実現されています。しかし、今後AI技術がさらに発展し、実世界での管理・運用・制御といった幅広い領域に適用される可能性を考えた場合、人間とAIの「協調」がきわめて重要な課題となるでしょう。

さらに、AIが人間の知的判断を完全に代替する未来を想定する場合、重要なのはAIと人間の間の「コミュニケーション」です。この分野はまだ十分に研究が進んでいません。また、AI技術には、学習に用いたデータセットがカバーしない問題に対して精度が著しく低下するという本質的な課題があります。このため、AIにすべてを任せれば安全性が向上すると考えるのは危険です。

ただし、AIを情報の効率的活用のためのツールとして用いることは積極的に進めるべきです。人間が苦手な大量の



東北大学大学院工学研究科
技術社会システム専攻教授

たかはし ましと
高橋 信

1964年3月生まれ。1986年東北大学工学部原子核工学科卒業。1991年東北大学工学研究科原子工学専攻博士課程修了。1992年京都大学原子エネルギー研究所助手。1996年東北大学大学院工学研究科量子エネルギー工学専攻助手、2000年同准教授、2002年4月東北大学大学院工学研究科技術社会システム専攻准教授、2012年8月同教授、現在に至る。

原子力システムの安全、リスクコミュニケーション、ヒューマンインタフェース設計と評価、航空システムのヒューマンファクター、脳機能イメージングの工学応用に関する研究に従事。工学博士。

〈受賞〉

1995年計測自動制御学会賞論文賞、2000年ヒューマンインタフェース学会学術奨励賞。

〈所属団体・委員〉

- ・日本原子力学会、ヒューマンインタフェース学会、計測自動制御学会、人工知能学会、日本人間工学会などの会員
- ・技術研究組合制御システムセキュリティセンター 理事長
- ・ヒューマンインタフェース学会監事
- ・日本人間工学会東北支部長・理事
- ・日本原子力学会東北支部長
- ・青森県原子力政策懇話会委員
- ・新潟県原子力発電所の安全管理に関する技術委員会委員
- ・原子力規制庁原子炉安全専門審査会審査委員