ラウンドテーブルⅡ

「AI 技術文明の基層２： AI ネットワーク化が浸透する社会における法と経済」

報告１ Alexander J. WULF〔アレクサンダー・ヴルフ〕

（ベルリン SRH 国際マネジメント大学教授）

“Artificial Intelligence and Transparency - Legal, Ethical and Technological Considerations in the Light of the EU’s GDPR”

はじめに、本会議にお招きいただき、ありがとうございます。「AI と透明性」について、EU の GDPR（General Data Protection Regulation）に照らした法的・倫理的・技術的な考察をお話しすることができ、光栄に思います。

私の本日の話ですが、透明性の原則に注目していただくところから始めたいと思います。これは日本の総務省から出ている「AI 開発ガイドライン案」において定式化されています。この原則は、AI システムの開発者は、AI システムの入出力の検証可能性及びその判断の説明可能性に留意すべきというものです。透明性の原則は、AI アルゴリズムの責任ある利活用に当たり不可欠となるものです。この透明性の原則の解説においては、総務省は、本原則の対象となる AI システムとして、利用者及び第三者の生命、身体、自由、プライバシー、財産等に影響を及ぼす可能性のある AI システムが想定されると述べています。そして、開発者は、AI システムの判断の説明性と同様に、その入力及び出力の検証可能性に留意することが望ましいとしています。このように、当局たる総務省は、既に透明性が重要であるということをこのガイドライン案に定式化しています。

本日は、いくつかの論点についてお話ししたいと思いますが、それらは特に EU の GDPR に目を向けるものです。私はこれを世界史上最も現代的なデータ保護規則だと考えていますが、そこには二つの重要な原則があります。すなわち、AI アプリケーションに対する透明性の権利（15 条）及び自動化された意思決定に対する制限（22 条）であり、これらは既に存在し、発効しています。

ここでは法的な論点だけに留めたくないので、本質的に、なぜ AI システムの透明性が重要なのかということもお話ししたいと思います。本日既に「ブラックボックス」の話が出ています。それは、AI システムがどのように動いているのかを表す、ある種のシンボルです。プログラムをした人でさえ、それがどのように動いているのかわからないときがあるのです。人工知能（以下、AI）の技術的な基礎を簡単に見直しておきたいと思います。具体的には、機械学習のアプローチにおける、AI システムの予測（回帰）モデルという古典的アプローチと、ニューラルネットワーク及び深層学習アルゴリズムと

いう現代的アプローチです。お示ししたいのは、数年前から進展が顕著となってきた現代的なアプローチがブラックボックスを抱えているのに比べ、80 年代に発達した伝統的アプローチは、相対的には透明性が高いということです。

その上で、先の二つの法的な問題に注目していただき、私自身の経験的な調査の一部をお見せします。これは、今日の事業者が利用者に対して、法的な権利及び義務についてどのように知らせているのかを研究したものです。具体的には、利害関係者に対し、どのような未来を予測するのか、AI システムの顧客や利用者に権利や義務を知らせることがより重要になるのか、そして、個人情報保護上の一般的な論点について尋ねました。

結語として、私がなぜ透明性の原則が重要と考えるのかを示したいと思います。法の領域における、いくつかの慎重に扱うべき機械学習及び AI のアプリケーションの例を概観します。その上で、透明性の問題に対する技術的な解決策も概観していきます。これは、この問題について将来的に見出されるかもしれない技術的解決がどのようなものであるかについて、正しい方向を素描する第一歩になると見ています。

一般的な言及から始めますが、まず、機械学習及び AI は、最近になって全く新しいものとして開発されたものではありません。これらは、いずれも 20 世紀の後半にはあったものです。80 年代・90 年代には、AI への初期のアプローチが確立しました。当時は AI と呼ばれておらず、データマイニングやコンピュータ統計学と呼ばれていましたが、保険や金融といった領域で使われていました。この初期のアルゴリズムは、現在も有効に使われています。しかし、ここ 10 年間に沢山の技術的な発展があり、具体的には、より多くのデータが利用可能になり、より強力なデータ処理能力が利用可能になってきました。AI システムをプログラムしやすくするデータがクラウドにあり、増強された処理能力があり、より高度な統計的アルゴリズムがあり、より良いユーザーインタフェースがあるわけです。簡単な AI アプリケーションを作ろうと思えば、最近ではあまり難しいことではありません。

しかし、より速い機械学習と AI アルゴリズムを採用する実用的な理由もあります。基本的に、AI アルゴリズムというものは、予測の精度が人間を凌ぐことが多々あります。適切にプログラムされていれば、人間よりもバイアスのかかっていない決定を下せる場合があることもまたアドバンテージとなり得ますしょう。ひとたび AI システムがセットアップされれば、人間よりもはるかに高速に、何百万もの個人データを分析し、数秒で自動的に決定を下します。そして、AI アプリケーションは、安価でもあります。埋没費用や高いセットアップコストを含めても、一度配備してしまえば、極めて低い費用で運用することができます。

AI を含む最新又は将来のアプリケーションのほとんどは、機械学習や深層学習アルゴリズムに基づくものであり、今日お話ししたい問題、AI システムに関する主な課題の一つは、それらがブラックボックスの中で作動しているということ、意思決定に透明性が

なく、どのようにその決定に至ったのかがわからないということです。このことが、長期にわたり AI アプリケーションの採択を妨げる道徳的・倫理的・社会的・法的な問題を生み出しています。現代の技術、例えば原子力エネルギーや遺伝子工学などにも、制御を失うおそれが多分にあります。そのような、制御を失うおそれが十分に払拭できなければ、AI の発達や普及も妨げられることになるでしょう。先の講演でもお話があったように、AI が色々と有益な効果をもたらすとしても、透明性に対して納得のいく形のアプローチがなければ、それらは利用できません。ブラックボックスの中で行われる深層学習アプローチが多くの AI アプリケーションにおいてほとんど不可避のものになっているのは、深層学習のアルゴリズムが複雑で多次元のデータセットからパターンを見つけ出し、知識や理解を示すことができる最新のアルゴリズムであるからです。透明性についてお話をする前に、人工知能の技術的な基盤について簡単に見てみたいと思います。

本質的に、古典的な機械学習のアプローチとの区別をすることができます。古典的な機械学習のアプローチは、予測（回帰）モデル、スコアリング、ディシジョンツリーに基づいており、ほぼ線形で透明性があります。社会科学者、例えば私などは、経済学の研究に当たり、この線形の回帰モデルを、法の変化が企業や市民の行動にどう影響するのかを予測するために用いました。伝統的な機械学習のアプローチは、線形で透明性があるので、ニューラルネットワークと深層学習に基づく最新の機械学習のアプローチと同じ問題は生み出しません。最新の機械学習のアプローチは、本質的に不透明なアルゴリズムであるといわれています。

この図で簡単にご説明しましょう。これは、古典的な線形の回帰モデルがどう働くのかを極めて単純に描いたものです。例えば、企業が利用者にオンラインで信用取引を提供したいとします。この利用者が信用取引をするに値するか否かを判断するために、多くの変数、例えば年齢、性別、所得、デュレーション等を考慮します。これらの変数は、単純な線形の回帰モデルをプログラムするために用いられ、プログラマーは、債務不履行の可能性という出力（outcome）に対し、どの変数が用いられ、どう影響しているのかを、直接的に述べることができます。利用者は、新しくローンを組める可能性、その場合の債務不履行の可能性を計算することができ、銀行は、債務不履行の可能性が５％以上であれば信用取引に値しない等の、切り捨ての相場（rate）を定めることができます。このように、プログラマーは、回帰モデルがどの変数を用い、それがどう働くのかを明示することができ、それぞれの変数が計算に関わっている程度、出力にどう影響するのかを説明することも容易です。線形の回帰モデルについて要約しますと、わかりやすく、追跡可能な方法で計算されており、比較的、人間にとって理解し、翻訳しやすく、出力についても比較的コミュニケーションしやすく、このモデルに基づく意思決定も、必然的に比較的正当化しやすい、ということになります。

最新のニューラルネットワーク及び深層学習のアプローチにおいては、システムは当

然異なっています。これらのアプローチに透明性はありません。再び入力変数を見て頂きますと、先の図のように一つの出力変数を予測するために用いられるのではなく、たくさんの出力変数を予測するために用いられています。加えて、この第一のレベルで予測された出力変数、所謂ニューロンは、第二の層でさらなるニューロンを予測するために再び用いられ、それが次々と繰り返されます。さらに、入力変数にはそれぞれ無作為に異なる重み付けがなされており、そのため、無作為な変化により極めて複雑なネットワークが作り出されます。多層レイアウト及び入力変数への無作為の重み付けにより、出力変数を再度入力変数と関連付けることは容易ではなくなります。ニューラルネットワークの問題は、最適の予測能力を持つモデルを用いているとしても、無作為な重み付けと多層デザインにより、単純な回帰モデルに比べて、入力変数が用いられる範囲を示すことができない、ということです。複数のニューロンの層を加えることで、ニューラルネットワークは、データの中から捉えにくいパターンを見抜くことができますが、それらを翻訳することは非常に難しくなりました。そして、この膨大な層により複雑性が高まり、予測（出力）がデータと関連付けにくくなっています。そのため、計算はわかりにくく、結果についてもコミュニケーションしにくくなっています。AI による判断もまた正当化されにくくなります。では、これがなぜ問題になるのでしょうか。

今年、GDPR が施行された際に、利用中のオンラインサービスから通知を受け取った方も多いかもしれません。我々は皆、自分が使っていた多くのサービスにおけるデータに係るプライバシー、すなわちプライバシー規則が更新されたという通知を受け取りました。EU が、最新の先進的なデータ保護法である GDPR を設けたことによるものです。GDPR が今回のお話においてなぜ興味深いかというと、二つの重要な規定が盛り込まれているからです。すなわち、AI アプリケーションの透明性に対する権利に関する規定及び自動的な意思決定の制限に関する規定です。そもそも、欧州議会及び欧州理事会は、これら AI システムを法の内に組み入れる規則を設ける導入するために、次のような理由付けをしています。データ主体、つまり利用者や消費者に対し公正かつ透明性のある処理を確保するため、管理者（controller）はプロファイリングに当たり適切な数学的又は統計的な手続を用いるべきであると述べているのです。このことは、AI アプリケーションの多くに関係してきます。例えば、先に見たようなオンラインのクレジットアプリケーションや、AI システムが自動的に応募者に対して判定を行う求人アプリケーションにも当てはまります。実際に、私の学生の一人が、午前３時にアプリケーションに登録し、２時間後の午前５時に不合格の通知を受けたと言っていました。彼女は、このアプリケーションでの審査は AI が行っていたと推測してもいました。なぜ立法府（legislator）がこのような権利を設けたのかということがこのことから明らかとなります。不透明な AI アルゴリズムを採用しては欧州の立法府が示す要求を満たすことができないことを問題としているのです。

では、実際の条文を見てみましょう。まずは、AI アプリケーションの透明性に対する権利です。GDPR の 15 条に規定されています。【訳者注：同条 1 項においては】データ主体は管理者から以下の情報を取得する権利を有すると規定し、当該情報として(a)取扱いの目的…(h)プロファイリングを含む自動化された意思決定の存在、当該意思決定に含まれる論理（the logic involved）に関する意味のある情報（meaningful information about the logic involved）等が掲げられています【訳者注：実際には、同条第 1 項は、「The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information」とした上で、(a)から(h)までの「information」（情報）を、「the right to obtain」（取得する権利）の対象としてではなく、「the right to … access」（アクセスする権利）の対象として列挙している。】。EU の GDPR は、AI を備えた製品やサービスを用いる全ての利用者に対し、その AI アルゴリズムを採用している者が誰しも、当該 AI アルゴリズムがどのように働き、自動化された意思決定においてどのように機能しているのかについて、情報を取得する権利を与えています。そして企業は、AI の意思決定に含まれている論理回路について説明しなければなりません。これは極めて壮大で進歩的なアプローチです。なぜなら、今や企業が直面する問題が見えているからです。企業は素晴らしく新しい製品やサービスを提供するために AI システムを採用していますが、そのアルゴリズムは不透明です。しかし、欧州の立法府は、これら企業に、AI を備えた製品やサービスの利用者に対して、AI アルゴリズムがどのように働くのかを透明性のある方法で説明することを求めています。将来的に、これが実際にどのように働き実施されるかはわかりませんが、現実に実施された AI に対する最初の重い規則であると考えています。

次いで、二つ目の制限について見てみましょう。自動化された個人に関する意思決定（Automated Individual Decision-Making）の制限です。ここでは、EU の立法府は、【訳者注：GDPR22 条 1 項においては】データ主体は、自らに関し法的効果を生じさせ、又は自らに関し同様の重大な影響を及ぼす、プロファイリングを含む自動化された取扱いのみに基づく決定の対象とされない権利を有すると規定されています。ここで立法府は、一歩踏み込んでさえいるわけです。すなわち、立法府は、人間、消費者及び企業の顧客が AI による自動化された意思決定の対象とされるとともに、その決定が重大かつネガティブな効果を利用者に与えた場合には、いつでもその AI による決定を行った企業に勤める別の人間の手で再検討することを要求しているのです。

求人アプリケーションの例を考えてみてください。先に、午前 3 時に求職した学生の例を出しました。職というものは人間の存在に関わる重大事ですので、求人を却下されることは、正に重大な決定といえます。さらに、もし彼女が仕事を得ていたら問題にはなりません。ネガティブな結果がない以上、AI システムが「OK です。貴女に職を与え

ます」と言っていたら問題はないのです。しかし、自動化された方法で不合格とされたとなれば、EU の立法府は、AI システムがこういった決定を単独で行うことはできず、重要なネガティブな決定があったときには人間が AI の決定を再検討しなければならないと求めるわけです。再検討によって、EU の立法府は、ブラックボックスを確認するよう定めているだけではありません。適切に再検討し、AI システムとは異なる決定をするための十分な裁量を持つよう定めているのです。

さて、法について私が申し上げたいことは以上です。そして、先に、技術的な問題について、AI アプリケーションは、透明性のない最新のニューラルネットワークを用いていると申し上げました。しかし、EU の立法府は、このことについて、企業は、透明性のあるコミュニケーションをするため、こういったアルゴリズムについて説明しなければならないと要求しているわけです。また、重大でネガティブな決定については人間が再検討しなければならず、AI が単独でそのような決定をしてはならないと求めているのです。

ここからは、いくつかの経験的な調査をお見せしたいと思います。独国におけるオンラインでの情報開示についてのステイクホルダーに対する質的調査です。これは、企業が消費者に対しどのように権利と義務の情報を提供しているのか、透明性があるのかを調べるものです。Amazon、Apple、iTunes、Facebook 等のサービスに加入した人はその際に何ページにも及ぶ文書が出てくることを誰しもが知っていることでしょう。法関係者を除き、それらをきちんと読む人はいないと思います。私の調査は、企業に対する権利及び義務を本当に理解させるため、いかにしてこの長い約定について消費者に透明性のある形で情報提供するのかということについてのものです。その調査の一部をお見せしますが、これは、独国のステイクホルダー（消費者、消費者団体、裁判官、政治家、企業、法律家）に対し、将来的に、消費者に情報を提供しなければならないということが一層重要になってくると考えるのかについて質問をしたものです。

私たちの質問は、「AI を備え自律的に作動する製品の普及によって、情報に係る義務が将来更に重要になると思うか」というものです。「情報に係る義務」とは、立法府の企業に対する要求における権利及び義務について消費者・顧客に情報を提供すべきことをいうものです。ここから、全体の過半数にあっては消費者の権利及び義務についての法的な情報公開が更に重要になると述べていることがわかります。この統計について一つ強調しますと、これは質的調査であり、割合は、各々の話題がインタビュー中にどれくらいの頻度で話題に出たのかを数えたもので、複数回答が許容されています。21%とあるのは、インタビュー全体の 21%で話題に出たということです。全体の 21%が「情報公開は、更に重要になる。AI 製品が一層複雑になってきているので、消費者にそれがどのように働いているのかを伝えなくてはならないからだ」といっています。20%が「情報公開は、一層重要になる。我々が AI システムを更に用いるようになれば、プライバ

シー問題も更に重要になってくるからだ」といっています。11%が「更に重要になる。AI システムを更に使うようになれば、消費者の権利が更に重要になるからだ」と答えています。責任問題に言及した人もいます。自動運転車が壁にぶつかった場合等を考えているのでしょう。情報公開の重要性は現在と同程度に留まるだろうと言った人は、少数でした。重要性が低下すると答えた者は、ごく少数でした。

個々の意見について見てみましょう。一人の裁判官は、「AI を使った製品の複雑さによって、情報公開がますます重要になるだろう」と言っています。その裁判官によれば、「そのことは想像できる。情報についての責任がある理由は、既存の構造的な不均衡を釣り合わせるためだからだ。それは、知識を持つ者にあっては、契約相手が知識を持っていない場合においては、情報を提供することを義務付けられているということである。このことは、消費者が製品を使えるために情報を必要とする場合にはいつでも当てはまる」ということです。つまり、私たちが AI を使った製品やサービスを使う場合、企業は、消費者の権利について一層多くの情報を提供しなければならない。消費者は、もはや製品のことを簡単には理解できないからであると言っているわけです。私がトヨタの車を買ったならば、それがどう働くのかについては何とかして知ることができます。しかし、自動運転車については、知ることができない。よって、企業は、このような製品を使用する際の消費者の権利や義務について、これまで以上に情報を提供する必要があり、そこに透明性をもたせる必要があるということです。

もう一人の裁判官は、「顧客の権利及び義務という点でより重要になろう」と言っています。その裁判官によれば、「それは、必要になろう。なぜなら、顧客は、製品について何が起こり得て、誰に責任があり、実際のところ責任とは何を意味するのかを承知していないからである。例えば、車は、誰かが傷害を受けるかもしれない。そのとき私が責任を負うことになるのか。消費者に対し AI を使った製品の透明性を確保するため、情報に係る義務は大幅に強化しなければならないこととなろう」ということです。

それから、消費者団体は、「個人データの保護の観点から、一層重要になる」と言っています。「しかし、ここで更に重要な点は、データ収集、データ保管、データ処理、データ譲渡についての透明性を確保しなければならないということである。どんな情報が収集されているのかについて、様々な段階において、消費者は知らされなければならない。それが中心的な問題である」ということです。

透明性がこれ以上重要になることはないという意見も見てみましょう。実際に、透明性について批判的な意見を示した企業があります。その企業によれば、「『あなたの携帯電話は、あなたがどこに居て、どこに移動し、どこで運転をしているのかいつでも知っている』と十年前に言われたとしたら、怖かったことであろう。AI を使った製品やサービスも同じで、それらは知らぬ間に我々の生活に入ってくるため、今は大きな恐れがあっても、すぐに私たちは慣れて忘れてしまう。携帯電話、フィットネストラッカー等他

の製品に起きたことと同じだ」ということです。他の企業は、「楽観視している。消費者に対する更なる情報公開は必要ない。我々は消費者に AI を使った商品について情報を提供できる AI アルゴリズムを開発することであろう」と言っています。問題は、それ自身によって解決されるというわけです。

次の質問は、「パーソナルデータの保護の問題について、企業は消費者に十分な情報を提供していると思いますか」というものです。AI の文脈でなぜこの質問が重要なのかといえば、AI システムは膨大なデータを必要とする機械学習アルゴリズムに基づくものであるからです。ある消費者団体は、「問題の核となるのは、単に AI アルゴリズムの透明性だけではない。AI アルゴリズムがどのようなデータを使っているのかということも重要である」と言っています。

私たちは、現時点でパーソナルデータの保護について消費者は十分に情報を提供されているのかを質問しました。御覧のとおり、多くの回答者は、「十分ではなく、透明性もない」と言っています。消費者は十分な情報を提供されておらず、情報を得たとしても理解できなかろうということです。例えば、Facebook のデータ保護方針、あるいは他のデータ保護方針を読もうとしますと、大量の情報がそこにあったとしても、普通の消費者は、その会社が自分のデータをどう扱っているのかがわからない。法的義務は果たしているのかもしれないが、その技術において実際のところデータがどう扱われているのかはわからないということです。

「情報は十分であるが、透明性がない」という人もいます。30 頁にわたってデータ保護に関する情報が開示されているのはよいが、それを理解はしていないといっているわけです。弁護士等は、「十分であり、透明性もある」と言っています。ある弁護士は、「私にはそこに書いてあることを明確に理解できる」と答えました。私たちは、消費者だけでなく、弁護士事務所や専門家たちとも話をしています。「情報は十分ではないが、透明性はある」と言った人も、少数ながらいます。「一般化することはできない。データ保護についての情報の開示が上手くいっているところもあれば、下手なところもある」と言った人もいます。ここには様々な意見があります。インタビューから具体的な回答を取り上げ、更に明確にしていくことにしましょう。

ある消費者団体は、「パーソナルデータの保護についての情報は、十分ではなく、透明性もない」と言っています。その消費者団体によれば、「プライバシーに関するお知らせが法的に必要となることを満たしていないということではなく、結局、消費者は自分のデータがどう扱われているのかがわからないということである。その記述が抽象的に過ぎ、又は得られる情報があまりにも乏しいからである」ということです。また、別の回答者は、「上手くやっているところもあれば、上手くいっていないところもある。パーソナルデータの保護方針を一般化したり、全般的に最適化されているとすることは、明らかに事実ではない」と言っています。最後に、別の回答者は、「情報は、十分であり、透

明性もある」と言っています。これは弁護士の意見ですが「必要な情報は全てそこに書かれており、読めば理解できる。要するに消費者はわざわざそれを読んで理解しようとしないだけである」ということです。他にも「消費者には情報を得る機会が確かに与えられている。指摘したい点は、消費者が自らのデータがどう扱われるのか、AI システムがどう機能するのかを気にしていないということである」という意見もあります。

では、なぜこういった問題が重要なのでしょうか。先立って、AI システムの技術的な基礎を見直し、ニューラルネットワークのプログラムは、従来の AI や統計的アルゴリズムに比べて不透明であると述べました。次いで、法的に必要となることを確認し、この問題についての経験的な調査を見てきました。本日のお話の最後に、最も慎重に扱うべき法の分野における機械学習及び AI のアプリケーションの例を見ていきたいと思います。

米国においては、裁判官が保釈、量刑及び仮釈放について判断するのを支援するために、AI が一層使われるようになってきています。再犯リスクを計算したり、被告が公判でどう主張するのか等を評価するのに使われるようになってきたということです。私は、ベルリン地区裁判所において商事裁判の手伝いをしていますが、普通、裁判は年間予算の中心的な部分を担うものとは見られておらず、一方で、裁判官はきわめて多くの事案に対応しなければなりません。その中で、早期に保釈するのか否かを判断する場合に AI アルゴリズムを使うわけです。米国においては、裁判官にこういったプログラムを販売する会社があります。犯罪者の特定のデータを提供し、再犯リスクを計算してくれるのです。再犯リスクのある犯罪者であれば、裁判官が早く釈放するということはないでしょう。どちらかというと、長きにわたり釈放せずに留めておくと考えられます。

なぜこのことが問題なのかといえば、まず、こういったプログラムは民間企業が作成しており、どのようなプログラミングになっているのかについて、オープンソースではなく、完全なブラックボックスになっている。裁判官も、弁護人も、一般市民も、その仕組みがわからない。そのような AI アプリケーションが人生を大きく変えるような判断に使われるわけです。例えば、誰かが公判に出るべきか、量刑はどうなるのか、保釈を許して良いのか、早く釈放してもよいのかといったことを裁判官が判断するための支援に使われています。さらに、こちらに示す例は、ウィスコンシン州とエリック・ルーミスとの間の裁判ですが、この AI システムを使った判断により、被告人は早く保釈してもらえなかった。他の人は保釈してもらえたのに、なぜ自分はしてもらえなかったのかについて、その理由を透明化できないと被告人は主張しているわけです。さらに、弁護士又は法律家は、このような用途に AI システムを使うということは、法的原則、例えば法の下における平等、一般市民に対する説明責任といったものを損ねる可能性があると指摘しています。こういった AI アルゴリズムは、企業が開発しているためにブラックボックス化されていることに加え、そうしたプログラムに対し現状では法的な認証

や承認を得る必要がないためです。

あまり負の側面ばかりを言及したくはないのですが、私自身が裁判に出席して考えるのは、裁判官に対する期待は大きく、バランスのとれた意見が必要だということです。純粋に商業的な領域であれば、適切な条件下で AI システムを活用できるかもしれません。しかし、刑事裁判手続で、人間が問題とされるとき、単なる商事事件ではないという場合はどうでしょうか。そのような場合についても、既に米国ではそれが実現されているわけです。ベルリンの州の高官は、こういったシステムが独国ですぐに導入されることはない、と懐疑的ではありますが。

最後に、どのような解決策があり得るか、ということを示したいと思います。現状を見ますと、AI のアルゴリズムに、特に倫理的な側面、例えば法律の領域において透明性が欠けているということは、大きな問題です。しかし、この状況は、間もなく変わるかもしれません。様々な科学者が解決策を開発しようとしています。機械学習や人工知能を、ブラックボックスではなく、「ガラスボックス」化しようとしているわけです。これまで、人工知能や機械学習のシステムは、主に営利目的の団体によって開発されてきました。よって透明性というのは、それほど重視されてこなかった。さらに、それに関連する法律もほとんどありませんでした。開発を推し進めたのが営利目的であることは、もちろんそれでよいのですが、今初めて、日本で開発ガイドラインが示され、GDPR がヨーロッパで発効する中で、法的、倫理的及び道義的な問題に関する議論が必要だと考えます。

そのような問題を意識し、科学者及び監視者の方では、AI システムの透明性を確保する措置を開発しようとしています。例えば、独国の科学研究機関の一つであるフラウンホーファー研究機構のサイトにおいては、自分で用意した写真を AI システムに認識させることができます。これは私自身の写真ですが、AI システムは、これはおそらく「スーツ」、それから「ネクタイ」であろうという結果を出しました。このような結果がどのようなアルゴリズムから生じたのかということです。かつて、この AI システムは完全にブラックボックスでしたが、フラウンホーファー研究機構は新たなアルゴリズムを開発しています。少なくともそれを辿れば、この画像認識のプロセスの中で、画像の中のどのような要素を使って分類したのかということがわかります。例えば、肩の部分、首の部分、こういった要素を使って、このような分類に至ったということです。私の頭の部分等は、それほど使われていないことがわかります。さらに、明確な誤りがわかることもあります。例えば、私の目の部分を使って「スーツ」と分類していますが、これは、間違いです。このような解決策を、量刑の判断や司法の現場にそのまま応用できるのかはわかりませんが、これは一つの事例といえます。私たち科学者、こういったシンポジウムに集まる専門家の皆さんが、適切なガイドライン、適切な原則を提示すれば、科学者の方は儀を変えて、AI の透明性を確保する技術を開発してくれると考えます。そうす

れば、こういったシステムも、より幅広く、社会で認められていくと思います。

これまで見てきたように、機械学習やAIアプリケーションは、道義的・倫理的機微に触れる分野、例えば医学それから刑事司法、さらに金融といった分野でも、ますます重要になってきています。法的、倫理的及び道徳的価値のためには、その AI アプリケーションの用途、そしてその基盤にあるアルゴリズムの機能に関し透明な情報開示が必要となります。例えば、法的な要求としては、GDPR の例を示しました。さらに、透明性というのは、将来の AI や機械学習アプリケーションの開発及び運用において重要な要件となってきます。特に、自己プログラミングできるアルゴリズム、ニューラルネットワークや深層学習が使われるような場合です。現在のところ、透明性が未だ十分ではありません、しかしながら将来においては、この透明性の確保のために様々なアプローチが開始されていくと考えています。

Artificial Intelligence and Transparency

Legal, Ethical and Technological Considerations in the Light of the EU's General Data Protection Regulation (GDPR)

Prof. Dr. iur. Dr. rer. pol. Alexander J. Wulf, MLB (WHU/BLS), MSc (LSE)
SRH University Berlin
e-mail: contact@alexanderwulf.eu

Draft AI R&D Guidelines by the Japanese Ministry of Internal Affairs and Communications et al.

Principle of Transparency

Developers should pay attention to the verifiability of inputs/outputs of AI systems and the explainability of their judgments.

- AI systems which are supposed to be subject to this principle are such ones that might affect the life, body, freedom, privacy, or property of users or third parties.
- It is desirable that developers pay attention to the verifiability of the inputs and outputs of AI systems as well as the explainability of the judgment of AI systems within a reasonable scope in light of the characteristics of the technologies to be adopted and their use, so as to obtain the understanding and trust of the society including users of AI systems.

Contents

Introduction (I)

- Machine Learning (ML) and Artificial Intelligence (AI) are no new developments, they date back to the second half of the 20th century.
- However, in the last decade technological developments have led to more data and more and more powerful data processing capabilities being available.
- While we currently observe increased adoption of these technologies, the future is likely to bring even more disruptive innovations in this field (e.g. self-driving cars, medical diagnosis, etc.).
- Practical reasons for the adoption of ML and AI technology: high accuracy, unbiasedness, fast, cheap.

See Finlay, 2017

Introduction (II)

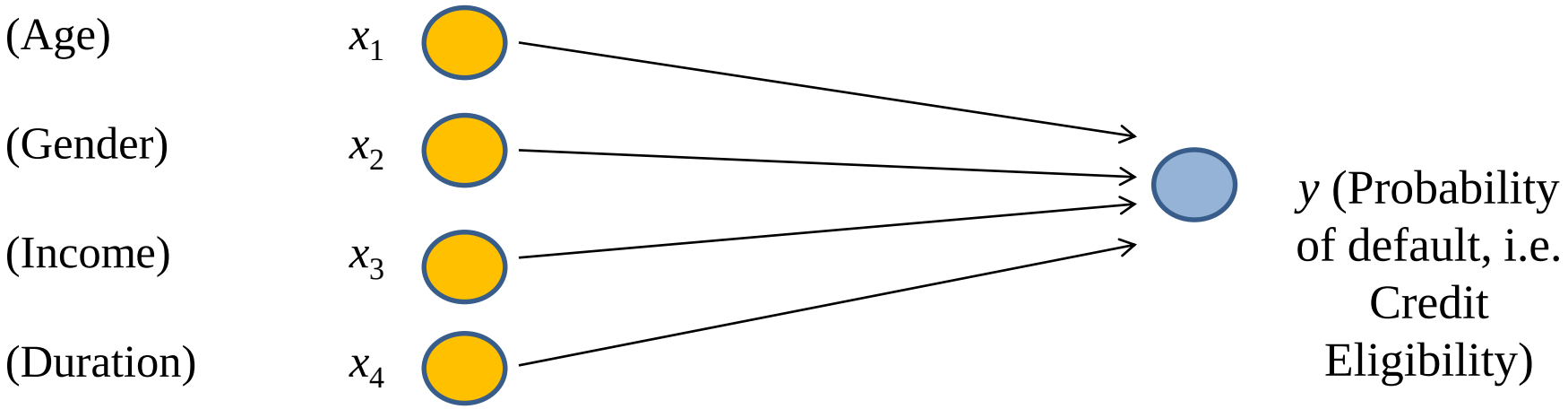
- Many current and future applications involving AI are based on machine learning (deep learning) algorithms.
- A major challenge pertaining to these AI systems is that they operate in a “Black-Box” and their decision making is thus intransparent. This creates moral, ethical, social and legal problems which may hamper the adoption of AI applications in the long run, e.g. “fear of losing control”.
- To employ deep learning approaches is nevertheless almost unavoidable for many AI applications, as these algorithms can find patterns (i.e. knowledge and understanding) in complex and multidimensional datasets.
- In what follows, I review the technological and legal problems pertaining to the transparency of AI algorithms.

See Holzinger, 2018

Machine Learning: The Technological Foundations of Artificial Intelligence

- Classical ML approaches are based on predictive (regression) models, scorecards and decision trees. They are (mostly) linear and transparent.
- Modern ML approaches are based on neural networks and deep learning approaches. They are non-linear and intransparent.

Classical Approaches: Predictive (Regression) Modelling

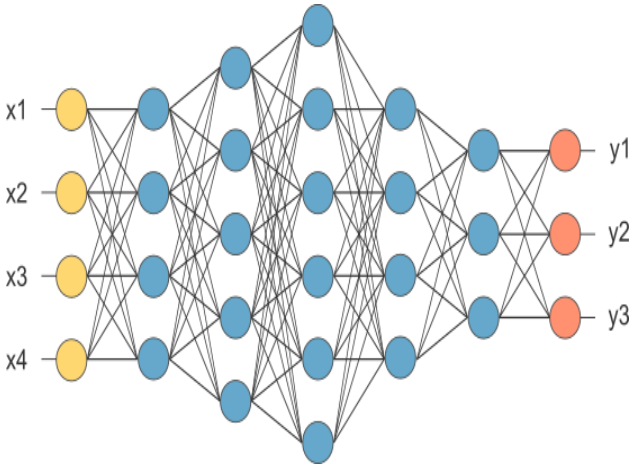


Transparency of Basic Predictive (Regression):

- Calculated in a comprehensible and traceable manner
- Relatively easy to understand and interpret by humans
- Results can be (relatively) easily communicated and decisions made on their basis thus (relatively) easily justified

See Finlay, 2017, Goodfellow, 2016.

Modern Approaches: Neural Networks and Deep Learning



The Deep Learning Process:

- Single neurons are similar to a basic predictive (regression) models.
- Neurons attach different weight to the input data, leading to different predictions (outputs).
- Outputs from the first layer are used as the inputs for the second layer, and so forth. Again, inputs on a given layer are the same, but due to different weighting lead to different predictions (outputs). The resulting neural network can get very complex.
- Machine learning refers to the process in which the weights are continuously adjusted and checked against known data until prediction is as far as possible optimized.

Transparency of Neural Networks and Deep Learning:

- By adding several layers of neurons, neural networks can detect subtle patterns in the data.
- A large number of layers leads to greater complexity. Predictions (outputs) cannot easily be linked to input data.
- Calculations are not comprehensible, results can thus not be easily communicated. Decisions based on predictions cannot be easily justified.

See Goodfellow, 2016., Finlay, 2017

The EU's General Data Protection Regulation (GDPR): Artificial Intelligence and Transparency

- The General Data Protection Regulation (GDPR) is the EU's main law on data protection.
- When the GDPR came into force on 25 May 2018, it led to worldwide discussion and controversy in business, politics and society due to its new and far reaching regulations.
- The GDPR introduces:
  - A Right to transparency for Artificial Intelligence applications (see next slides).
  - A Restriction of automated individual decision-making (see next slides).
  - Recital 71: "In order to ensure fair and transparent processing in respect of the data subject, (…) the controller should use appropriate mathematical or statistical procedures for the profiling (…)."
- Examples: AI decision-making regarding online credit or job applications.
- Problem: As described above, many AI applications (Deep Learning) cannot explain themselves. They are intransparent by design.

European Parliament and of Council, 2016

A Right to Transparency for AI Applications

Article 15 of the GDPR: Right of Access by the Data Subject

1. The data subject shall have the right to obtain from the controller (…) the following information: (…)

(a) the purposes of the processing (…)

(h) the existence of automated decision-making, including profiling, and, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

(…)

European Parliament and of Council, 2016

A Restriction of Automated Individual Decision-Making

Article 22 of the GDPR: Automated Individual Decision-Making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

(…)

European Parliament and of Council, 2016

Empirical Research on How Different Groups of Stakeholders View Online Information Obligations

Consumers

Consumer Organizations

Judges

The Stakeholders of Online Disclosures in Germany

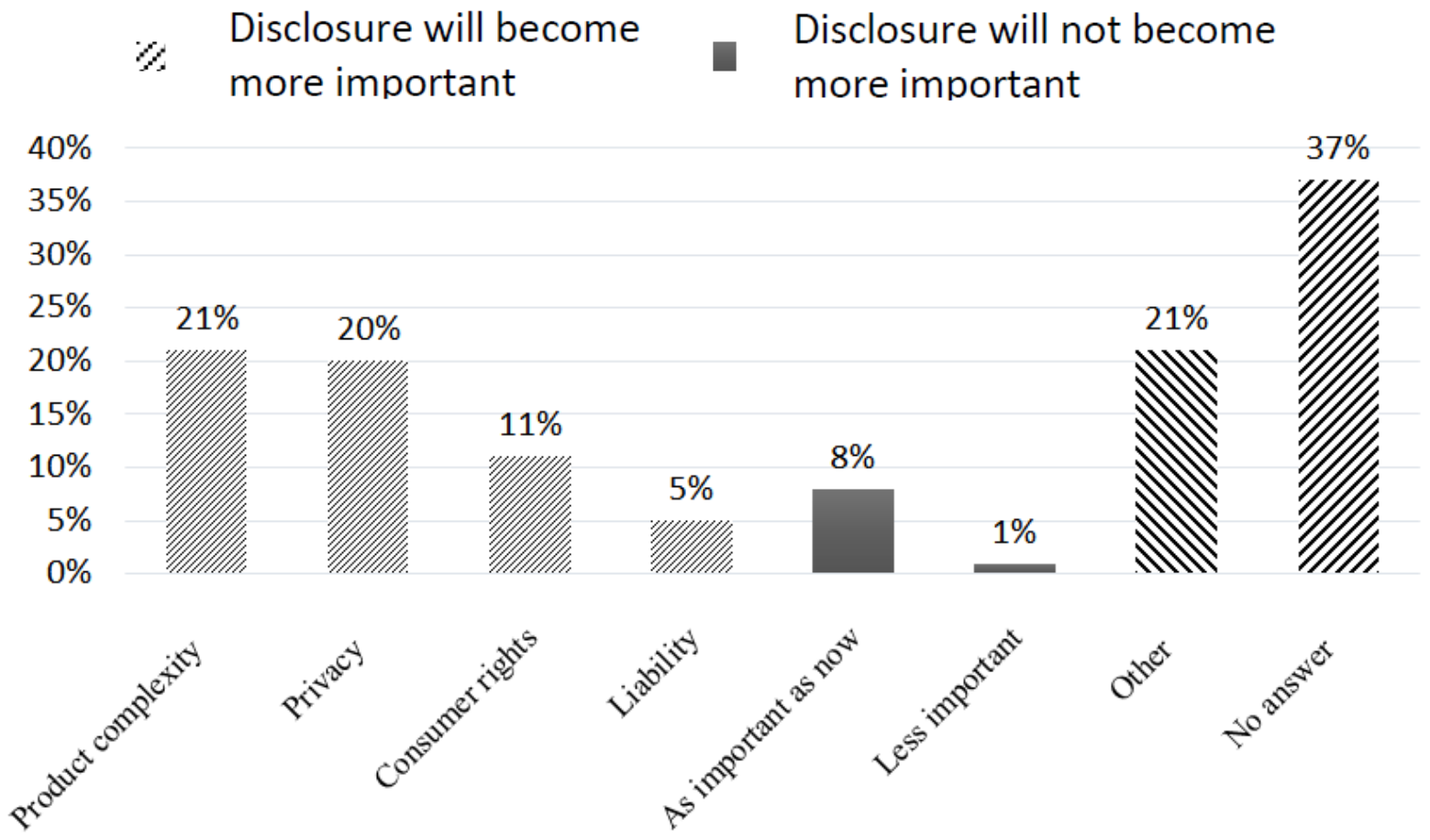
Politicians

Businesses

Lawyers

European Parliament and of Council, 2016

Do you believe information obligations will become more important in the future thanks to the proliferation of autonomously operating products powered by AI? (Summary Statistics)



Seizov and Wulf, forthcoming

Do you believe information obligations will become more important in the future thanks to the proliferation of autonomously operating products powered by AI? (Qualitative Results I)

- More Important: Product Complexity

  „Yes, I could imagine that. At least, I think it's plausible, since the reason for the information obligations is often to balance out an existing structural imbalance. That is, a person who has knowledge that the contract partner does not have is obliged to provide [that] information. And with that in mind I think that is always the case when the consumer needs information to be able to use a product at all." (Judge)

- More Important: Customer Rights and Obligations

  “Yes, it will be necessary, because customers are not aware of what could happen or who is liable and what liability means in practice in regard to the product. For instance, with cars, that somebody could be injured. And am I the one who is then responsible, I am liable in such a case? I think the information obligations will have to be substantially increased." (Judge)

- More Important: Personal Data Protection

  “But what's much, much more important here is the fact that we [must] have transparency in regard to data collection, data storage, data processing and the passing on of data, that the consumer is informed about what is collected and that on the different levels. As I see it this is a central issue." (Consumer Organization)

Seizov and Wulf, forthcoming

Do you believe information obligations will become more important in the future thanks to the proliferation of autonomously operating products powered by AI? (Qualitative Results II)

- As important as now

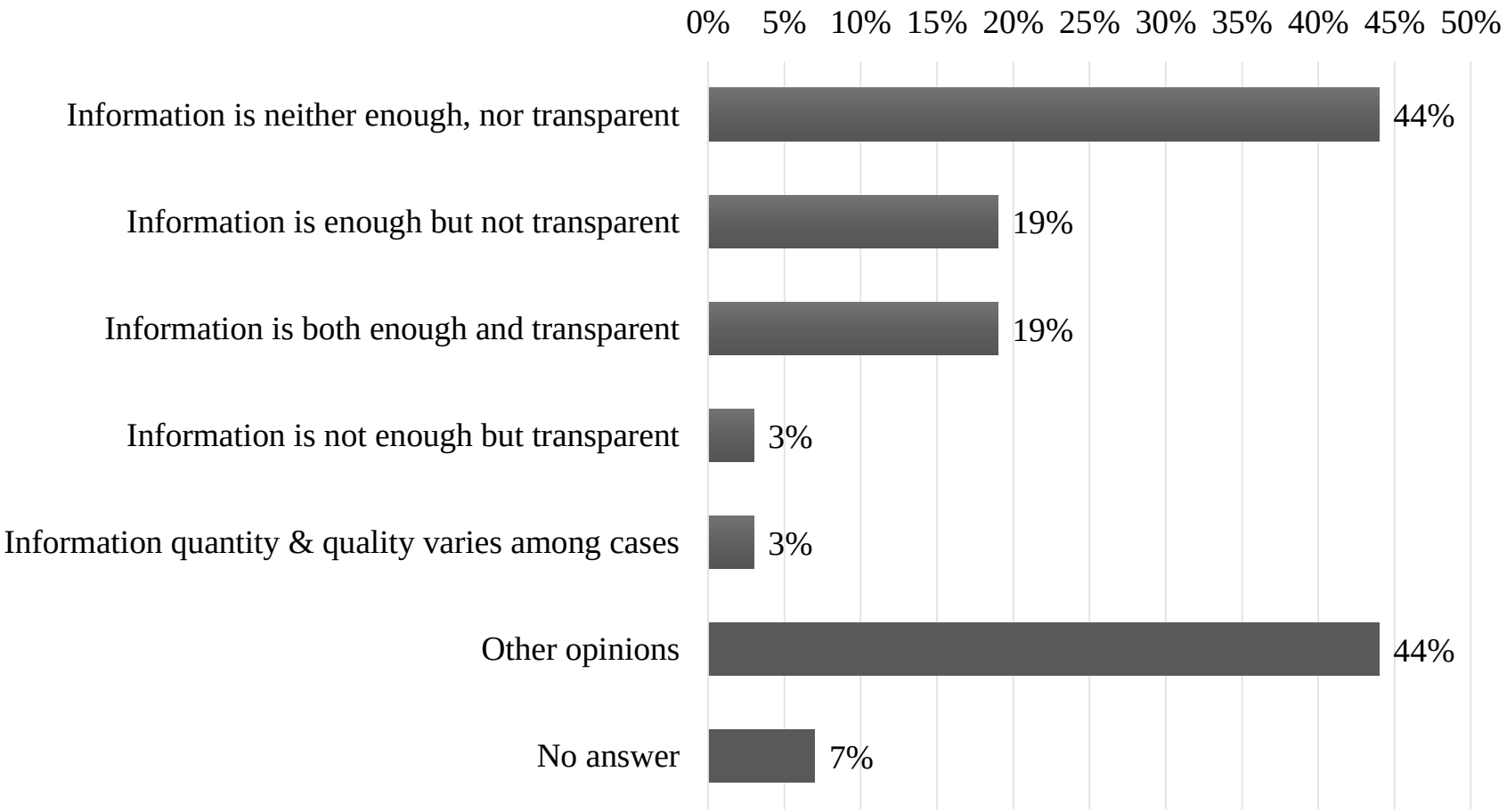
  “On the side of the consumer I expect that that will not have much of an influence on marketing or PR because some will say, "That's bad!" and others won't notice it at all. If somebody had said 10 years ago, "Your cell phone always knows where you are!", there would have been an enormous outcry. But I haven't noticed anything of the kind in the last 10 years. And now my phone is recording my pulse rate, knows exactly where I am and constantly hears everything that's going on. So ... and all that has happened because of customer profiling. And that being as it is I assume that people won't notice it at all, as long as the tools are slowly introduced to them. Ultimately you see it in the fitness tracker industry, too." (Business)

- Other

  "Haha. Well, I'm an optimist as regards the future and so I would think that in the future artificial intelligence will help us to better comply with our information obligations. That's the bottom line." (Business)

Seizov and Wulf, forthcoming

Do you think businesses inform consumers sufficiently about matters of personal data protection? (Summary Statistics)



Seizov and Wulf, forthcoming

Do you think businesses inform consumers sufficiently about matters of personal data protection? (Qualitative Results)

- Information is neither enough, nor transparent

  "While we are not saying that [privacy notices] fall short of the legal requirements, we believe that in the end consumers still do not know what happens with their data. Either because the description is too abstract [or] because information ends up being quite scarce" (Consumer Organization)

- Information is enough but not transparent

  "Some do it well, others not so well. To generalize and say [privacy policies] are totally optimal across the board is definitely not the case." (Anonymous Participants)

- Information is both enough and transparent

  "The information is all there, if one wanted to go through it"
  "The consumer definitely has the opportunity to get informed"
  (Varying Anonymous Participants)

Seizov and Wulf, forthcoming

An Example of a Sensitive Application for ML and AI in the field of Law

- In the USA, AI applications are increasingly used to assist judges when making decisions about bail, sentencing and parole. They supposedly help a judge better assess, e.g. a criminals reoffending risk, how likely the defendant is going to appear for trial, etc.
- AI recommendations, such as risks scores for reoffenders directly influence important trial outcomes, such as sentencing terms (see e.g. *Wisconsin v. Loomis*).
- Lawyers fear that these applications may erode important legal principles, such as the rule of law, equality before the law, public accountability, etc. because these algorithms are programmed by private companies and operate as “Black Boxes” with no need for official certification or approval.

Tashea, 2017

Technical Solutions to the Transparency of Machine Learning and Artificial Intelligence (I)

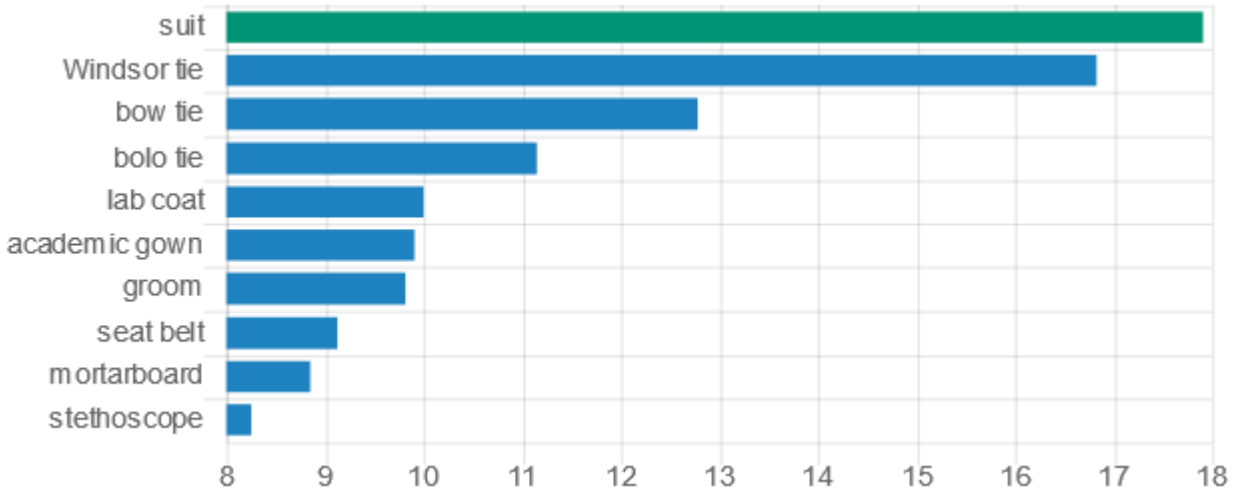
- Modern ML applications which are based on neural networks and deep learning approaches are non-linear and intransparent.
- New technical solutions to the transparency of ML and AI try to get from a “Black Box” to “Glass Box” setting via post-hoc interpretation approaches
- Post-hoc Interpretability does not necessarily require a precise (mathematical) definition of how a model works, but rather relates to textual, visual or conceptual explanations.

See Litpon, 2016

Technical Solutions to the Transparency of Machine Learning and Artificial Intelligence (II)

Original

Heatmap for the Predicted Class



The image was classified as suit with a classification score of 15.71

https://lrpserver.hhi.fraunhofer.de/

Conclusions

- Machine Learning and Artificial Intelligence applications are becoming more and more relevant, also in fields that are sensitive to moral and ethical considerations (e.g. medical applications, criminal justice, financial applications etc.).
- Legal, ethical and moral values require transparent disclosures about the functioning of an AI applications' underlying algorithms.
- Transparency is thus a key requirement for the future development of AI and ML applications, particularly where self-programming algorithms (Neural Networks and Deep Learning) are involved.

References

- The Conference Toward AI Network Society. *Draft AI R&D Guidelines for International Discussions*. Tokyo: Ministry of International Affairs and Communications, 2017.
- Finlay, Steven. *Artificial Intelligence and Machine Learning for Business. A No-Nonsense Guide to Data Driven Technologies*. Great Britain: Relativistic, 2017.
- Goodfellow, Ian et al. *Deep Learning*. Cambridge: MIT Press, 2017.
- Holzinger, Andreas. *Interpretierbare KI. Neue Methoden zeigen Entscheidungswege künstlicher Intelligenz auf*. Hannover: Heise Medien, 2018.
- Lapuschkin, Sebastian et al. "The LRP Toolbox for Artificial Neural Networks ", *Journal of Machine Learning Research*, 17, 2016 , 1-5.
- Seizov, O. and Wulf, A. J., "Can't Live with Them, Can't Live Without Them: How Different Stakeholder Groups in Germany View Online Information Obligations", *German Law Journal*, forthcoming.
- Tashea, Jason. *Courts Are Using AI to Sentence Criminals. That Must Stop Now.* San Francisco: Weired, 2017.