2026年度

一般選抜　前期日程

グローバルマネジメント学部
グローバルマネジメント学科

小論文

（　80分　）

注意事項

1　試験開始の合図があるまで，この問題冊子は開いてはいけません。

2　問題冊子は８ページあります。解答用紙は２枚，下書き用紙は１枚あります。

3　試験開始の合図後，まず，問題冊子，解答用紙の落丁，乱丁，印刷不鮮明等がないか確認し，気付いた場合は，手を挙げて監督者に知らせてください。

4　試験開始後，受験番号，氏名を解答用紙の所定欄（解答用紙１枚につき，受験番号２箇所，氏名１箇所）に記入してください。

5　試験開始後は，原則として，試験が終了し退出許可が出るまで退出できません。

6　解答は，解答用紙の指定された箇所に，横書きで記入してください。
解答用紙にアルファベット，算用数字を記入する場合，１マスに２文字ずつ入れてください（ただし，字数が奇数の場合は，末尾の１文字は１マスに入れてください）。

7　解答用紙は，持ち帰らないでください。

8　試験終了後，問題冊子および下書き用紙は持ち帰ってください。

問題

AIやロボットを規制する法のあり方について論じた次の文章を読んで，以下の問いに答えなさい。

中世ヨーロッパには「動物裁判」と呼ばれる社会制度があったことが知られている。たとえば疫病が流行したとき，その原因と疑われた動物――典型的にはネズミやネコ――が数匹捕らえられ，法廷で有罪を宣告されて処刑されたり，不幸にも赤ん坊を蹴り殺してしまった豚が追放刑に処されたりしたというのだ。さてわれわれはこの制度を正当なもの，意味のあるものだと考えるだろうか。考えないとすれば，それはなぜなのだろうか。

おそらくただちに現われる反応は科学的な合理性がないとか，このようなことをしても動物が一定の行為（たとえばネズミによるペストの媒介）を止める見込みがない以上，意味はないというものだろう。法制度には行為指導性，つまり特定の行為を人々に行なわせようとしたり，逆にさせないようにする性格がある。典型的には一定の行為が処罰されるだろうという予測をもとにして，そのような行為への関与を事前に避けることが，人間にはできるわけだ。だが動物にはこのような予期能力がないので，違反行為を事後に処罰することを通じて学習させるしかない。犬猫のしつけとはこのように事後処罰に頼ったシステムであり，だからこそベンサム(注1)は何が禁止されているのかを個々人が理解することのできないようなイングランドの判例法システムを「犬法」と罵倒したのだった。

逆に言えば，同じヒトに属していても一定の理性や判断能力を持たない存在を法の指示に服従させることはできないし，彼を法廷に引き出して有罪を宣告するようなことも，動物裁判の例と同じように無意味だということになる。もちろん心神喪失者が刑事的に処罰されないことを定めた刑法39条1項が，ここで思い出されることだろう。①法を通じた事前規制は，結果を予期し自らの行動をコントロールすることのできる自律的主体に対してしか意味を持たないと，とりあえず確認しておこう。

19世紀半ば，カール・フォン・サヴィニー(注2)が確立したドイツ民法学が世界を〈人〉と〈物〉とに区分し，理性的主体である〈人〉が合意に基づいて相互に形成する関係として契約をとらえようとしたこと，あるいはその前提にあったカントの哲学・倫理学もまた同様の観点に立っているし，日本の現在の民法もその延長線上にあることに注意しよう。理性に基づいて自律的に行為することのできる存在だからこそ人間は〈人〉として根元的な自由を認められるのであり，そのような資格を持たない存在はすべて，その生命の有無にかかわらず，〈物〉として〈人〉の意思に従属させられることになるわけだ。

動物の権利をめぐる生命倫理学的な問いも，基本的にはこの構図を継承している。たとえばある種の動物はヒトに近い知性や自律性を持つので限定的な権利能力を認められるべきだという主張は，しばしば見受けられるだろう。逆に，胎児や重度心身障害者は意識と受苦能力を欠くので生命権のような人権保障の対象にならないという倫理学者ピーター・シンガーの主張も見受けられる。その具体的な主張内容はさまざまであっても，これらはいずれも意思と自律性を持つ〈人〉を高く，それらを欠く存在を低く位置付けるスロープの存在を前提にしている点ではほぼ違いがない。どこにその両者を区切る境界線

があるのか，あるいは境界線は単一でなくいくつかの段階から構成されているのか――たとえば犬猫には能動的な政治参加の権利はなくとも理由なく傷付けられない権利は認められるべきかもしれない――，その境界線と生物種としてのヒトの境界が一致しているのかといった論点をめぐってどれだけ激しい論争を繰り広げたとしても，われわれの社会を構成する〈人〉の特権性が，そこでは常に前提されているのである。

ここで問題は，いま新たにわれわれの世界に生まれつつある存在としてのAIやロボットが，この構図のなかでどのように位置付けられるかという点にある。もちろんわれわれは，これまでの生命倫理学と同様の構図に立ったうえで，境界線をめぐる議論に基づいてその地位について考えることができる。〈人〉の条件を理性や判断能力に求めれば，AIやロボットも人間と同じ高みに昇るべきだということになるかもしれない。痛みや苦痛を感じる能力だと考えれば，動物より低い位置付けになることも考えられるだろう。だがそのような議論の構図は，AIやロボットの本質的なあり方を正確に反映しているのだろうか。AIやロボットは，まだ十分に技術が発展していない現時点では動物のようなものであり，成長するにつれてヒトと同じような存在へと移行していくだけのものなのだろうか。

少し遡って考えよう。さきほど法制度には行為指導性があり，処罰が予告された行為を人々は選択しなくなるだろうと述べた。だがそこでいう行為指導性がそれ自体として実在するのならば，処罰の予告を伴う必要はないのではないだろうか。たとえば「殺人は悪いことであり，禁止される」とだけ述べればよく，それに一定の制裁――現行法では死刑または無期もしくは5年以上の拘禁刑（刑法199条）(注3)――を結び付ける必要はないのではないだろうか。

もちろんこの問いへの答はごく明らかであり，人々が規範に従いそこねるという点にあるだろう。われわれの多くは，あえて刑法の条文によって示されることがなくとも，殺人が悪い行為だということは十分に理解している。だとしても一定の理由からその悪をあえて選択したり（あいつだけは生かしておけない），あるいは自らの行動を理性的にコントロールしそこねることによって（ついカッとなって），悪いと知っているはずの行為に手を染めてしまうわけだ。そこでわれわれは，制裁を予告することで合理的な判断者にとってのバランスを変えてしまおうとする一方（処罰されることを考えれば，相手を殺すのは割に合わない），激情に駆られた行為者に対しては実際に処罰を科すことで，他の人々への戒めにしようとすることになる。

いずれにせよ法は，それが人間の判断や行為に直接的に介入できないこと，判断や行為の条件を操作することで間接的に機能することを狙うしかないということを前提としている。ここでは再び，法は外形的な行為を規制する合法性の次元においてのみ働くものであり，内心にある道徳性を左右することはできないというカントの指摘が想起されるだろう。だからこそ，物理的な条件を通じて判断・行為の可能性自体を事前に消去するアーキテクチャ(注4)の権力――たとえば宝石を入れた展示ケースに壊せない鍵を付ければ，見た人がそれを盗みたいと思うかどうかにかかわらず，盗めないようにできるだろう――が，法を超えるものとして注目を集めることにもなるはずだ。

このように考えたとき，②AIやロボットがヒトと大きく異なる点として，そのような従いそこねの

可能性に注目することができるだろう。まず，現在でも生産現場で活用されているようなロボットを考えよう。それらはあらかじめ定められたプログラムに沿って定められた動作を反復し続けるだろうし，故障や燃料切れといった物理的な障害の場合を除けば，それに失敗することもないだろう。プログラムはロボットの動作を直接的に規定するのであり，そこには判断も自律も，したがって従いそこねの問題も生じないように思われる。

まさに最近話題となっている自動運転車のように，高度な学習機能を備えたAIや，それによって制御されるロボットの場合にはどうだろうか。もちろんそこに，与えられたデータからAIが何を学習するかが予測しにくいとか，データ自体に偏りが含まれていればそれをAIが忠実に学習してしまうという問題があることはすでに指摘されている。

たとえば，イギリスの病院で研修医を選ぶ際にそれまでの判定結果を学習させたAIを使ったところ，結果的に女性が不利に扱われていたことが事後に発覚したという事例が伝えられている。これは，過去の女性差別的な社会において当時の病院関係者たちが意識的にか無意識にか女性を不利に扱っていたという経緯があり，その際の判断データを学習したので，そこに含まれていたバイアスがそのまま再現されてしまったという問題だと考えることができるだろう。

また，画像認識に基づいて被写体に関するキャプションを付けるAI――たとえばこれは自転車だとか，これは卒業式だという内容を画像自体から判断するというシステムが，黒人の写真を「ゴリラ」と判断してしまったという問題も報道されている。これは，開発者が学習のために利用したデータに含まれていた人間の顔写真が人種的に偏っており，白人の写真ばかりが用いられていたことの結果だと言われている。黒人の顔という正解データを見たことがないので，見たことのあるなかで一番近いもの――「ゴリラ」だと考えてしまったというわけだ。このような問題からは，不適切な結果の発生を防止するために学習データの適切性を担保することが実際の応用に向けた課題として指摘されることになるだろう。

2016年3月には，米マイクロソフト社がTwitter上で公開した会話AI「Tay」が，ユーザーとの会話を通じて人種差別や陰謀論をたちまち学習してしまった結果として不適切な発言を連発するようになり，半日あまりで緊急停止されるという事態も生じている。やはり同様に，AIによる不法行為や暴走への懸念などが語られる一つの契機にはなったのかもしれない。

だが注目すべきなのは，ここでたとえば人種差別を学習したAIがただちにそれを実行に移していることだろう。AIは差別発言をすることが適切だと思ったからそれを直接に行動へと反映させたのであり，そこにはヒトの場合であればしばしば生じるようなさまざまな配慮（この発言は社会的に許容されるか，TPOに合致しているか）やためらい（面倒くさい，眠い，疲れた）は存在しない。学習結果はAIの行動を直接的に規定しており，ヒトのような従いそこねの可能性はここにも存在しないということになるのではないだろうか。AIやロボットが我ら人間とは異なる「超人」的なあり方を実現するものだとすれば，それはたとえば理性，知識量，判断能力，情報処理速度といったもので人間を大きく凌駕するような「超知性」（superintelligence）だからではなく――あるいはそれに加えた別の問題とし

て――，③このようにわれわれとは根本的に異なった規範への反応構造を持っているからだと考えられる。

出典：大屋雄裕「AIにおける可謬性と可傷性」宇佐美誠編『AIで変わる法と社会―近未来を深く考えるために』（岩波書店，2020）45～50頁。ただし，出題にあたって一部改変した。

（注）

※1　ベンサム　Jeremy Bentham（1748～1832）。古典的功利主義を提唱したイギリスの哲学者，法思想家。

※2　カール・フォン・サヴィニー　Friedrich Carl von Savigny（1779～1861）。ドイツの法学者。

※3　刑法199条　現行法では，「人を殺した者は，死刑又は無期若しくは五年以上の拘禁刑に処する。」

※4　アーキテクチャ　ローレンス・レッシグが論じた規制手段のひとつ。物理的環境に手を加えることによって，人の行動のコントロールが試みられる。

問1　下線部①に関して，どうして「自律的主体」ではないものに対して法による事前規制が意味を持たないのか。120字以上170字以内で説明しなさい。

問2　下線部②に関して，AIやロボットの「従いそこねの可能性」の有無について，180字以上230字以内で説明しなさい。

問3　下線部③に関して，「われわれとは根本的に異なった規範への反応構造」を持つAIやロボットの判断や行動に対する規制のあり方について，本文での記述を踏まえ，あなたの考えを300字以上400字以内で述べなさい。