„Ethos" 36(2023) nr 4(144) 231-235

Damian SZCZĘCH

PLAN NA ŻYCIE W ŚWIECIE ZDOMINOWANYM PRZEZ SZTUCZNĄ INTELIGENCJĘ

Catriona Campbell jest psychologiem behawioralnym, przedsiębiorcą i doradcą biznesowym. Specjalizuje się w design thinking oraz tematach związanych z interakcją człowiek–komputer. Recenzowany tekst jest jej pierwszą książką. Cel autorki stanowi przedstawienie stosowanej przez nią metody dążenia do pożądanego efektu końcowego (metoda future-back) oraz pokazanie, jak tę metodę można wykorzystać do wykreowania pewnej społecznej rzeczywistości i stworzenia planu działań, który ma doprowadzić do jej realizacji.

Tekst składa się z krótkiej biografii, podziękowań, ośmiu rozdziałów i wniosków. Pierwszy rozdział, zatytułowny „Sleepwalking into Singularity" („Lunatykowanie w stronę Osobliwości", s. 1-15), rozpoczyna się od banalnego stwierdzenia, że sztuczna inteligencja (ang. Artificial Intelligence, AI) używana jest wszędzie. Autorka przedstawia rozróżnienie słabej (wąskiej) oraz silnej (ogólnej) AI oraz krótki opis realnych i możliwych czy spodziewanych zastosowań. Następnie omawia koncepcję Osobliwości, zgodnie z którą maszyny sterowane przez AI mają zyskać świadomość i przewyższyć ludzi we wszystkich aspektach, co może stanowić dla nich zagrożenie. Przytacza ogłoszony w roku 2015 *Open Letter on AI* („List otwarty dotyczący AI"), którego sygnatariusze nawołują do podjęcia badań nad społecznym wpływem tej technologii. Rozdział kończy się stwierdzeniem, że musimy zaakceptować nadejście Osobliwości lub czegoś do niej zbliżonego i zaplanować, jak z nią żyć. Rozdział drugi, „AI by Design and the Future-Back Methodology" („AI według projektu i metoda future-back", s. 17-25) rozpoczyna się od wskazania czterech obszarów, którym – według omawianego „Listu..." – należy poświęcić uwagę: (1) optymalizacja wpływu AI na ekonomię, (2) prawo i etyka, (3) bezpieczeństwo i weryfikacja, (4) kontrola. Zdaniem autorki należy skupić się na planowaniu działań wobec zagrożeń płynących ze strony technologii. Pomocna ma tu być metoda future-back, stosowana w marketingu i procesie myślenia projektowego (ang. design thinking), bazującego na tworzeniu rozwiązań opartych na zrozumieniu faktycznych potrzeb potencjalnych użytkowników. Metoda future-back polega na przyjęciu w punkcie wyjścia pożądanego przyszłego stanu rzeczy i z tej perspektywy opracowaniu strategii mającej doprowadzić do jego realizacji.

Rozważania rozdziału trzeciego, „Should We Be Afraid of the Current State of AI?" („Czy powinniśmy się obawiać AI w obecnym kształcie?", s. 27-49), otwiera

[1] C. C a m p b e l l, *AI by Design: A Plan for Living with Artificial Intelligence*, CRC Press, New York 2022, ss. 164.

teza, że obecnie jesteśmy świadkami swoistego „wyścigu zbrojeń", analogicznego do tego z czasów zimnej wojny. Celem tego wyścigu jest stworzenie silnej AI, czyli systemu przewyższającego ludzki intelekt pod każdym względem. Autorka omawia problemy wynikające z zastosowania AI w wojnie cybernetycznej i konwencjonalnej. Rozdział kończy optymistyczne stwierdzenie, że obecnie mamy jeszcze wpływ na rozwój AI i możemy tym rozwojem pokierować według naszego planu. W rozdziale czwartym, zatytułowanym „Current State of AI Governance & Regulation" („Obecny stan zarządzeń i regulacji dotyczących AI", s. 51-62), autorka wymienia szereg instytucji zajmujących się przygotowaniem zaleceń odnośnie do rozwoju AI. Zauważa przy tym, że liderzy rozwoju technologicznego – USA i Chiny – praktycznie nie ponoszą konsekwencji za ich nieprzestrzeganie. Wskazuje także na ogromną lukę w stanie wiedzy, występującą między organami tworzącymi regulacje a korporacjami tworzącymi technologie, bo to te ostatnie mają rzeczywistą wiedzę na temat powstających rozwiązań technicznych. Niestety autorka nie podaje rozwiązania, jak tę lukę można by zniwelować. Problem jest poważny także z tego względu, że zdaniem autorki przedsiębiorstwa – nastawione na zysk – nie są godne zaufania jeśli chodzi o wprowadzanie jakichś form samoograniczenia. Przy tworzeniu zasad i regulacji rozwoju AI konieczna jest wobec tego współpraca przedsiębiorstw z rządami. Niewiele jest w książce wskazówek, jak współpraca ta miałaby wyglądać, a to właśnie od jej formy i zasad zależy jej skuteczność przy opracowywaniu regulacji badań i wdrożeń AI oraz ewentualnych sankcji za ich złamanie. Rozwiązanie proponowane przez autorkę przypomina nieco teorię umowy społecznej, ponieważ zakłada, że możliwe będzie wypracowanie korzystnego porozumienia respektowanego przez wszystkich dla dobra wszystkich. Autorka pomija natomiast rolę organizacji pozarządowych (ang. non-governmental organizations, NGOs) w procesie nadzoru i tworzenia regulacji dotyczących AI. To poważne przeoczenie, gdyż pozostawia całość władzy w rękach rządów i korporacji, których członkowie są powiązani biznesowo i politycznie – często się zdarza, że wysocy rangą urzędnicy państwowi po zakończeniu kadencji pracują w biznesie, a specjaliści ze świata biznesu są zatrudniani w organizacjach rządowych. Z jednej strony pozwala to w jakiś sposób wypełnić wspomnianą wyżej lukę w wiedzy, ale z drugiej strony rodzi poważny konflikt interesów.

W rozdziale piątym, „Current State of AI Ethics" („Obecny stan etyki AI", s. 63-75), autorka opisuje słynne prawa robotów, sformułowane przez Isaaca Asimova: (1) Robot nie może skrzywdzić człowieka ani przez zaniechanie działania dopuścić, aby człowiek doznał krzywdy; (2) Robot musi być posłuszny rozkazom człowieka, chyba że stoją one w sprzeczności z pierwszym prawem; (3) Robot musi chronić samego siebie, o ile tylko nie stoi to w sprzeczności z pierwszym lub drugim prawem. Prawa te trudno jednakże stosować – twierdzi Campbell – ze względu na ich dużą ogólność. Dla ilustracji przytacza klasyczny w etyce dylemat wagonika, odnosząc go do samochodów autonomicznych. Zachęca wobec tego do bardziej elastycznego, kontekstowego podejścia, które określa jako „podejście usamodzielniające" (ang. empowerment approach) (por. s. 65). Zakłada ono, że roboty mają podejmować decyzje samodzielnie w oparciu o kontekst danej sytuacji, a nie o sztywne reguły. Najczęściej postulowane zasady to: transparentność, brak dyskryminacji oraz prawo do prywatności (por. s. 67). Proponowane przez Campbell podejście rodzi oczywiście wiele problemów, poczynając od tego, jak uczyć roboty takiej „moralnej samodzielności", aż po kwestię moralnej, ale i prawnej odpowiedzialności

za ich decyzje. Rodzi się też pytanie, jaki status miałyby mieć takie „usamodzielnione" roboty. Czy należałoby je traktować jako podmioty moralne? Konsekwencje postulowanego podejścia nie są przez autorkę analizowane. Autorka zauważa podobieństwo między etyką AI a etyką lekarską, ale podkreśla, że regulacje dotyczące AI są na bardzo wczesnym etapie rozwoju, podobnie jak sama AI.

W rozdziale szóstym, „Options for Our Future with AI" („Możliwe wersje naszej przyszłości z AI", s. 77-111), Campbell przedstawia w formie opowiadań pięć scenariuszy. Mają one pomóc nam uchwycić wersje przyszłości, do których nie chcemy dążyć. Nie bierze natomiast pod uwagę możliwości powstrzymania rozwoju AI – uważa to za nierealne. Pierwsza historia przedstawia dzień w Londynie, gdy rozwój AI pozostawiono samemu sobie: kwitnie cyberprzestępczość, ataki dronów są codziennością, technologiczne bezrobocie wzrasta, a społeczeństwo powoli się rozpada. Druga historia opisuje sytuację, w której rozwinięta w wysokim stopniu AI utworzyła sztuczny świat, w którym żyją ludzie (jest to wersja hipotezy symulacji Bostroma[2]). Trzecia historia jest kolejną wizją pesymistyczną – AI zaczęła pustoszyć planetę, a pozostali przy życiu ludzie uciekają do założonej przez Elona Muska kolonii na Marsie. Czwarta opisuje osadę zamieszkaną przez neo-luddystów, którzy wyrzekają się wykorzystania silnej AI oraz są odłączeni od publicznego Internetu i postawieni poza społeczeństwem. Ostatnia historia to opowiadanie o spotkaniu dwojga ludzi, którzy poddali się szeregowi zabiegów, by przekształcić się w cyborgi. Przedstawione scenariusze są dystopijne. Aby uniknąć realizacji któregoś z nich, konieczne jest regulowanie kierunków rozwoju i zastosowań AI. Warunek takiej regulacji stanowi wypracowanie pożądanej wizji przyszłości, a wtedy – właśnie metodą future-back – będzie można określić sposoby jej realizacji.

Rozdział siódmy, zatytułowany „Creating a Roadmap – a Plan for Living with Artificial Intelligence" („Tworzenie planu działania – projekt życia z AI", s. 113-137), zawiera główne postulaty autorki. W formie wezwania do działania przedstawia cele i ramy czasowe, w których powinny one zostać osiągnięte. Cele krótkoterminowe (2022-2025) obejmują między innymi stworzenie wspólnego modelu (ang. framework) zasad i regulacji (przy czym należy tu uwzględnić możliwie wiele różnych punktów widzenia), zwiększenie poziomu wiedzy technicznej prawodawców oraz uczynienie AI bardziej sprawiedliwą poprzez usuwanie uprzedzeń (ang. bias) ze zbiorów danych, na których opiera się jej działanie. Cele średnioterminowe (2025-2028) to: określenie zasad mających zapobiec nadużyciom i chroniących ludzi przed krzywdą wyrządzoną przy pomocy AI oraz stworzenie międzynarodowego porozumienia analogicznego do Traktatu o Nierozprzestrzenianiu Broni Jądrowej (ang. Nuclear Non-Proliferation Treaty, NPT). Jako cele długoterminowe (2025-2030) Campbell wskazuje: prowadzenie edukacji pozwalającej wszystkim lepiej zrozumieć AI i z nią współpracować, przekwalifikowanie pracowników oraz zarządzanie ekonomicznymi skutkami wprowadzenia AI na rynek pracy. Autorka nie wyznacza środków do osiągnięcia tych celów ani nie wskazuje, kto ma prowadzić i nadzorować edukację prawodawców oraz obywateli.

[2] Hipoteza: zaawansowana cywilizacja mogłaby stworzyć symulację (tj. komputerowe odwzorowanie rzeczywistości) obejmującą nasze umysły oraz otaczający nas świat (podobnie jak w filmowej trylogii *Matrix*). Główny argument polega na tym, że jeśli żyjemy wewnątrz takiej symulacji, to nie poznajemy „prawdziwego" świata. Ponadto nie jesteśmy w stanie zweryfikować czy żyjemy w symulacji, czy też nie (por. N. B o s t r o m, *Are You Living in a Computer Simulation*?, „Philosophical Quarterly" 53(2003) nr 211, s. 243-255.

Rozdział ósmy, „A New Hope for Artificial Intelligence" („Nowa nadzieja dla AI", s. 139-146), zawiera przemyślenia autorki dotyczące pożądanej wersji przyszłości. Campbell twierdzi, że odpowiednie zarządzanie AI może przynieść ludzkości ogromne korzyści. Pojawienie się Osobliwości może oznaczać nadejście nowego renesansu – ludzie będą mieć znacznie więcej wolnego czasu, dzięki czemu będą mogli rozwijać swoje zainteresowania, a ludzkość jako całość będzie dysponować znacznie większymi zasobami intelektualnymi, co znacząco przyspieszy jej rozwój we wszystkich aspektach.

W części pod tytułem „Conclusion" („Wnioski", s. 147-148) autorka wskazuje, że AI jest jednocześnie największym wyzwaniem dla ludzkości oraz jej największą nadzieją. Jest „boska" (ang. godlike) i tak rozległa, że zrozumienie jej ogromu przekracza możliwości pojedynczego człowieka. Dzięki niej zyskamy możliwość decydowania, kim chcemy być (por. s. 148), a wobec tego musimy zaplanować, jak chcemy żyć w świecie z AI.

Książka ma charakter popularnonaukowy, ale jest interesująca ze względu na to, że autorka znajduje się w gronie osób mających wpływ na realnie powstające rozwiązania techniczne. Campbell podejmuje tematy pojawiające się często w popkulturze, stosuje język potoczny i powołuje się na opinie „technologicznych celebrytów" pokroju Elona Muska (por. s. 11-15). Układ treści jest nieco chaotyczny. Podziękowania oraz pierwsze dwa rozdziały de facto stanowią właściwy wstęp. Struktura książki zostaje zarysowana dopiero w drugim rozdziale. Zasadnicze rozważania znajdujemy w rozdziale siódmym – zawiera on najważniejsze tezy i postulaty. Rozdziały nie są symetryczne: najdłuższy ma trzydzieści siedem stron, a najkrótszy jedenaście. Pierwszą częścią po spisie treści jest krótka i niezbyt informatywna biografia. Nie dowiemy się z niej, czym konkretnie zajmuje się autorka. Z treści książki wynika, że zawodowo jest doradcą biznesowym, specjalizującym się w design thinking, nie zostało to jednak sformułowane explicite. W tekście pojawia się także marketingowy żargon (np. testy A/B – por. s. 18) oraz akronimy instytucji (np. RNIB, RNID – por. s. 20), które nie zostały wyjaśnione, co utrudnia czytanie tekstu. Na stronie sześćdziesiątej pierwszej pojawia się błąd rzeczowy – przytoczono kazus procesu Sokratesa, słynnego filozofa i mentora Platona, a w następnym zdaniu stwierdzono, że to Platon został skazany na śmierć poprzez otrucie. Oczywiście skazanym był Sokrates, a Platon dożył spokojnej starości.

Autorka nie jest deterministką techniczną. Przyszłość pełni w jej rozważaniach rolę swoistej zasłony niewiedzy – trzeba podejmować decyzje, nie znając faktycznego przebiegu rozwoju AI, ale mając na uwadze pożądany scenariusz przyszłości. Campbell twierdzi, że rozwoju nie da się zatrzymać, ale jesteśmy w stanie nim kierować poprzez regulacje wypracowane przez kooperację rządów, korporacji oraz instytucji pozarządowych (NGO, tj. wszelkiego rodzaju fundacji i stowarzyszeń). Jest to dość utopijne podejście. Zakłada uczciwość wszystkich stron oraz możliwość stworzenia rozwiązań sprawiedliwych i korzystnych dla wszystkich. Wydaje się nie zauważać znaczącego konfliktu interesów w obszarze regulacji dotyczących AI. Ponadto zgoda na poziomie postulatów nie gwarantuje zgody na poziomie środków realizacji – o środkach i sposobach wypracowania owych regulacji autorka nie mówi zbyt wiele, a przecież zastosowanie środków też niesie ze sobą konsekwencje. Co więcej, w rozważaniach widoczny jest „zachodni" punkt widzenia, zakładający, że rządy demokratyczne będą zainteresowane rozwojem AI dla wspólnego dobra. Campbell w rozważaniach ignoruje istnienie niedemokratycznych czy wręcz auto-

kratycznych rządów, które też inwestują duże środki w rozwój AI i raczej nie kierują się w swoich działaniach wspólnym dobrem.

Zamysł autorki dobrze wpisuje się w koncepcję budowy tak zwanej odpowiedzialnej AI (ang. responsible AI), choć sformułowanie to nie pojawia się w tekście. Takie postulaty zwykle marginalizują rolę filozofii. Podobnie jest i tutaj: Campbell podejmuje rozważania na istotne filozoficznie tematy, pomija jednak rolę samej filozofii w wypracowywaniu rozwiązań. Traktuje etykę jako inspirację i punkt odniesienia przy projektowaniu regulacji, ale nie jako narzędzie do rozstrzygania problematycznych kwestii. Heurystyczna rola filozofii została całkowicie zignorowana. Sama autorka zaznacza udział filozofów w procesie wypracowania rozwiązań wewnątrz korporacji oraz w konsultacjach rządowych (por. s. 68, s. 71), udział ten jednak zazwyczaj ogranicza się do formułowania ogólnikowych zaleceń i wskazówek. Jest to bardzo niekorzystne dla wszystkich stron. Książkę można odczytać jako wezwanie pod adresem filozofów, by włączyli się aktywnie w działania dotyczące regulacji AI. Przeważnie rzadko zajmują się bieżącymi problemami, a obecnie na świecie dzieją się rzeczy wymagające podejmowania decyzji już teraz. Zwlekanie z wypracowaniem konkretnych stanowisk może doprowadzić do przeoczenia momentu, w którym jeszcze mamy wpływ na to, co się dzieje. Bez konkretnych stanowisk i związanych z nimi postulatów filozofia zostaje sprowadzona co najwyżej do roli komentatora. Staje się jednym spośród wielu źródeł opinii.

Systematyczny namysł nad techniką i konsekwencjami jej rozwoju, nawet związanego z AI, ekscytuje na ogół jedynie wąskie grono pasjonatów. Społeczeństwo interesuje się raczej spektakularnymi wydarzeniami, zwłaszcza katastrofami (czy możliwościami katastrofy). Dlatego też książki stanowiące próby zainteresowania szerszego grona taką tematyką zasługują na uznanie. Autorka jest osobą „nietechniczną", toteż stosowane przez nią opisy zawierają niewiele informatycznego żargonu, który mógłby odstraszyć czytelników. Książka stanowi ciekawą propozycję dla wszystkich zainteresowanych tematyką AI oraz nowych technologii – zawiera ogólny przegląd obecnego stanu regulacji oraz zachęca do podjęcia rozważań co do przyszłości człowieka i jego miejsca w świecie zdominowanym przez technologię. Jest również bardzo aktualna ze względu na opracowywane właśnie przez Parlament Europejski regulacje dotyczące zastosowań AI (tzw. AI Act), które uważane są za kluczowe, ponieważ będą wzorcem dla rozwiązań w innych częściach świata.

Kontakt: Wydział Filozofii, Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Racławickie 14, 20-950 Lublin
E-mail: szczech.dam@gmail.com