

Sztuczna inteligencja – ludzkie rozterki

„Nie mogę zaprojektować sprawiedliwego systemu SI” – stwierdza przyszły inżynier SI podczas seminarium z etyki SI. „Dlaczego?” – pyta wykładowca. – „Bo nie ma jednej definicji sprawiedliwości” – pada odpowiedź. – „To prawda, że nie ma. Jaka definicja sprawiedliwości jest ważna dla twoich interesariuszy i przyszłych użytkowników systemu?” – „Nie pytałem.”



Wojciech Bednaruk

od ponad 20 lat zajmuje się technologiami edukacyjnymi. Wdrażał systemy zarządzania rozwojem w Kanadzie, Europie Środkowo-Wschodniej i Afryce, tworzył cyfrowe programy szkoleniowe. Wykładowca etyki sztucznej inteligencji na Polsko-Japońskiej Akademii Technik Komputerowych, członek Sekcji AWSI przy PTI. Wydał pierwszą w Polsce książkę o etyce SI dla dzieci. Autor i prowadzący seminarium Etyka SI.



Ten dialog ujawnia klasyczny dylemat inżyniera SI poszukującego idealnego, uniwersalnego rozwiązania etycznego. Stwierdzenie „nie mogę zaprojektować sprawiedliwego systemu SI” sugeruje przekonanie, że system albo jest w pełni sprawiedliwy, albo wcale. Takie binarne myślenie pomija istotną prawdę – sprawiedliwość nie jest stanem absolutnym, lecz spektrum możliwości i kompromisów.

Co więcej, inżynier SI, poszukując idealnej definicji sprawiedliwości, wpada w pułapkę perfekcjonizmu technicznego. To prowadzi do paraliżu decyzyjnego – skoro nie można osiągnąć perfekcji, lepiej nic nie robić. Tymczasem w rzeczywistości społecznej często lepsze jest dobre rozwiązanie wdrożone dziś niż idealne, którego nigdy nie uda się osiągnąć.

Odpowiedź „nie pytałem” na pytanie o definicję sprawiedliwości ważną dla interesariuszy ujawnia zakorzenione przekonanie, że projektowanie systemów SI jest zadaniem czysto technicznym.

Tymczasem współczesne systemy SI to złożone ekosystemy społeczno-techniczne, w których technologia jest tylko jednym z elementów. Sukces takiego systemu zależy od zrozumienia kontekstu społecznego, kulturowego i organizacyjnego, w którym będzie funkcjonować. Dlatego kluczowe jest włączenie interesariuszy już na etapie definiowania problemu.

Gdyby projektant systemów SI najpierw zapytał przyszłych użytkowników o ich rozumienie sprawiedliwości, mógłby odkryć, że ich potrzeby i oczekiwania są bardziej konkretne i osiągalne niż abstrakcyjna „idealna sprawiedliwość”. To nie tylko oszczędziłoby zasoby, lecz przede wszystkim zwiększyłoby szanse na stworzenie systemu, który rzeczywiście służy swoim użytkownikom. W przypadku SI niezwykle ważne jest przejście od modelu „genialnego samotnego inżyniera” do podejścia bazującego na współpracy i dialogu z interesariuszami.



W etycznym projektowaniu SI nie chodzi o odkrycie uniwersalnych prawd, ale o uważne wsłuchiwanie się w potrzeby i wartości konkretnych ludzi, których system ma wspierać.



Jak uczyć etyki SI?

Wraz z zespołem AWSI postanowiliśmy podjąć wyzwanie wprowadzenia innowacyjnego podejścia do nauczania etyki SI i przygotować cykl trzech seminariów dla studentów uczestniczących w kursie „Wyjaśnialne uczenie ma-

szynowe – xAI” na Uniwersytecie Warszawskim. Zamiast tradycyjnych wykładów zaprojektowałem serię doświadczeń edukacyjnych, ułatwiających przyszłym inżynierom SI nawiązanie bardziej osobistej więzi z zasadami etycznymi. SeminaRIA były uzupełnieniem kursu prowadzonego przez prof. Przemysława Biecką, wzbogacającym techniczną wiedzę studentów o wymiar etyczny i społeczny.

Najpierw studenci poznali podstawowe pojęcia etyczne i ramy godnej zaufania SI. Szczególnie pouczające było ćwiczenie z analizą przypadku SafetyNet – aplikacji dla turystów, która miała ostrzegać przed niebezpiecznymi dzielnicami zwiedzanych miast. To ćwiczenie pokazało, jak pozornie neutralne decyzje techniczne mogą prowadzić do stygmatyzacji całych społeczności zamieszkujących dzielnice określone przez system jako niebezpieczne. Studenci rozważali też poczucie braku autonomii na podstawie doświadczeń aresztowania Józefa K., bohatera ponadczasowej powieści „Proces” Franza Kafki.

Drugie seminarium skupiło się na kwestii przejrzystości i sprawiedliwości algorytmicznej. Przyszli twórcy modeli SI analizowali przypadek z filmu „Raport mniejszości”, który pomógł im zrozumieć konsekwencje predykcyjnych systemów SI dla wolności jednostki. Wykorzystanie literatury i filmu pomogło przekształcić abstrakcyjne pojęcia w namacalne doświadczenia. Poprzez filmowe i literackie opowieści oraz komiksowe narracje wizualne studenci badali złożone dylematy etyczne z wielu perspektyw. To podejście okazało się szczególnie wartościowe dla inżynierów pochodzących w większości z polskiej klasy średniej, którzy choć technicznie wykwalifikowani, mogą mieć ograniczone bezpośrednie doświadczenie związane z dyskryminacją czy utratą autonomii.

Przyszli projektanci systemów SI analizowali oczekiwania i konsekwencje funkcjonowania systemu SI z wielu perspektyw – architekta systemu, lekarza, dyrektora szpitala i pacjenta. Student wcielający się w architekta systemu SI musiał uzasadnić pacjentowi decyzję algorytmu o obniżeniu priorytetu opieki i przesunięcia w czasie potrzebnego zabiegu. Takie ćwiczenia z odgrywaniem ról pomagają zrozumieć, jak abstrakcyjne zasady etyczne mogą wpłynąć na funkcjonowanie systemu, decydującego o życiu ludzi.

Badanie, dotyczące równowagi płci w edukacji IT, przeprowadzone z uczestnikami drugiego seminarium, pokazało, jak osobiste wartości wpływają na decyzje techniczne. Podczas gdy 75% respondentów (90% mężczyzn) poparło teoretycznie inicjatywę na rzecz różnorodności płci, 63% chciało, aby te zmiany nie miały praktycznego wpływu na obecne proporcje płci wśród przyszłych inżynierów SI. Ta rozbieżność między teoretycznym poparciem a praktycznym działaniem pokazała, jak osobiste interesy mogą nieświadomie kształtować decyzje etyczne.



Jak etycznie projektować?

Jedną z metodologii projektowania etycznego jest Value Sensitive Design, zaprezentowana uczestnikom trzeciego seminarium. Etyczne projektowanie SI zaczyna się od samoświadomości inżynierów – ich osobiste wartości działają jak filtry, poprzez które interpretują cele projektu i potrzeby interesariuszy. Presja biznesowa nie powinna automatycznie prowadzić do kompromisów etycznych, bo często uwzględnienie kwestii etycznych może zwiększyć zarówno korzyści społeczne, jak i wartość biznesową.

Sukces tej inicjatywy edukacyjnej sugeruje, że nauczanie etyki SI nie sprowadza się do przedstawienia ram i wytycznych. Każdy system sztucznej inteligencji jest w istocie manifestacją ludzkiej wyobraźni etycznej – naszej zdolności do przewidywania, jak technologia może i powinna wpływać na życie ludzi. Umiejętność dostrzegania i uwzględniania perspektywy interesariuszy w procesie projektowym będzie więc coraz istotniejsza.

Sokratejska maksyma „Poznaj samego siebie” niesie zaskakująco aktualny przekaz dla przyszłych twórców systemów sztucznej inteligencji, w których każda decyzja techniczna jest nieświadomie filtrowana przez pryzmat osobistych przekonań projektanta. Nieświadomy własnych uprzedzeń może nieumyślnie wprowadzić je do tworzonego systemu, powielając własne zniekształcenia poznawcze w kodzie i architekturze rozwiązania.

