



Jak Warszawa tworzyła etyczne ramy dla generatywnej AI w administracji publicznej

Czwarty czerwca 2024 roku. Pierwsze szkolenie z cyklu spotkań dla grupy roboczej, która ma przygotować etyczne ramy dla generatywnej AI w Urzędzie m.st. Warszawy. Pierwszy raz pracujemy bezpośrednio z narzędziem. Na ekranach ChatGPT. Przy stole dziewiętnaście osób z ośmiu biur urzędu: informatycy, prawnicy, specjaliści od marketingu, ludzie z HR, analitycy. Jedni otwierają narzędzie po raz pierwszy w życiu, inni już zdążyli je poznać na własną rękę. Prowadzący tłumaczy, jak formułować zapytania, jak oceniać jakość odpowiedzi, na co uważać.



Michał Kuszewski

główny specjalista ds. wspierania rozwoju innowacyjnej gospodarki w Biurze Rozwoju Gospodarczego Urzędu m.st. Warszawy, koordynator projektu „Kierunki odpowiedzialnego wykorzystywania generatywnej sztucznej inteligencji”. Łączy wykształcenie z zakresu psychologii i doświadczenie w projektowaniu UX z pracą w administracji publicznej. Członek Sekcji Aktualne Wyzwania Sztucznej Inteligencji przy PTI.



W połowie sesji ChatGPT przerywa pracę. Choć nasze komputery pracują bez zarzutu, a połączenie z siecią jest stabilne, okno czatu zastyga bez odpowiedzi. Jak się okazało, była to globalna awaria OpenAI.

Gdyby ktoś chciał wymyślić scenariusz szkoleniowy, który w trzydzieści sekund pokazuje, dlaczego nie można polegać na generatywnej AI jak na infrastrukturze, trudno było-

by o lepszy. Ta awaria nie spowolniła naszej pracy. Wzmocniła przekonanie, że podejście, które obraliśmy, ma sens.



Najpierw etyka, potem technologia

Projekt „Kierunki odpowiedzialnego wykorzystania generatywnej sztucznej inteligencji w Urzędzie m.st. War-

szawy” nie powstał w odpowiedzi na kryzys. Nikt nie popełnił spektakularnego błędu z użyciem AI, nie wybuchł skandal. Elementy sztucznej inteligencji od lat działały w narzędziach, z których urzędnicy korzystają na co dzień, często nawet o tym nie wiedząc. Ale gwałtowna popularyzacja jej generatywnej odmiany zmieniła sytuację: nagle każdy mógł wpisać pytanie w ChatGPT i dostać odpowiedź, która wyglądała jak napisana przez eksperta. Ludzie zaczęli testować, sprawdzać, podejmować pierwsze próby użycia w swoich zadaniach. Pytanie nie brzmiało „czy używać”, tylko „jak przygotować się na to, co nadchodzi”.

Decyzja, żeby zacząć od etyki zamiast od technologii, była świadoma i szła pod prąd. Większość organizacji robi odwrotnie: kupuje narzędzie, szkoli ludzi, a dopiero potem zastanawia się nad zasadami. My odwróciliśmy kolejność. To nie był konserwatyzm ani strach przed nowością. To było przekonanie, że administracja publiczna, która zarządza dwumilionowym miastem, nie może sobie pozwolić na eksperymenty bez ram.

Projekt wyrósł z programu „Generujemy Innowacje” w ramach Strategii #Warszawa2030 i zyskał patronat Ministerstwa Cyfryzacji. Inicjatywa nie przysła jednak z góry. Powstała oddolnie, z poziomu praktyka, który widział lukę i zaproponował, żeby ją wypełnić. Od początku założyliśmy, że powstaną dwa dokumenty: obszerny, szczegółowy zbiór wytycznych dla pracowników urzędu oraz krótki kodeks dla mieszkańców. Stolica musi dać obywatelom możliwość szybkiego wglądu w to, jak ich miasto podchodzi do sztucznej inteligencji.

Rozpoczęliśmy od wyboru partnera merytorycznego, szukając ludzi łączących umiejętności z zakresu facylitacji, etyki, technologii i projektowania usług. Wygrało konsorcjum Uniwersytetu SWPS, Centrum Etyki Technologii Instytutu Humanites i firmy EGO, specjalizującej się w *design thinking* dla sektora publicznego. Każdy z partnerów wnosił coś innego: SWPS dawał rygor akademicki, Centrum Etyki – normę i ramy, EGO – metodę pracy z grupą.

Grupa robocza to dziewiętnastu pracowników ośmiu biur urzędu: Biura Rozwoju Gospodarczego, Biura Informatyki, Biura Marketingu Miasta, Biura Strategii i Analiz, Biura Organizacji Urzędu, Biura Zarządzania Zasobami Ludzkimi, Stołecznego Centrum Bezpieczeństwa i Biura Prawnego. Wybór nie był przypadkowy, szukaliśmy jednostek, w których generatywna AI pojawi się najszybciej, gdyż zaproszenie wszystkich – przy czterdziestu jeden biurach w strukturze miasta – byłoby niewykonalne.

Rekrutacja przyniosła coś, czego nie planowaliśmy. Dyrektorzy typowali przedstawicieli, a ci okazali się zaskakująco różnorodni: obok entuzjastów technologii pojawili się sceptycy, którzy chcieli uczestniczyć właśnie dlatego, że mieli obawy. Ta mieszanka perspektyw okazała się najcenniejszym zasobem całego procesu. Entuzjaści widzieli możliwości, sceptycy widzieli pułapki. Razem widzieli więcej niż każda z grup z osobna.

Szkolenia i warsztaty

Półroczną pracę rozpoczęliśmy od cyklu szkoleń dotyczących samej technologii AI, po regulacje prawne i kwestie etyczne. Zależało nam, aby każdy członek zespołu najpierw zrozumiał, jak te narzędzia działają w praktyce, bo tylko wspólna płaszczyzna techniczna pozwalała nam rzetelnie oceniać ich konsekwencje. Te spotkania pełniły jednak jeszcze jedną, kluczową funkcję – budowały wspólny język.

” *Kiedy prawnik i informatyk zaczynają operować tymi samymi pojęciami, dyskusja o zasadach przestaje być starciem dwóch obcych żargonów, a staje się merytorycznym dialogiem.*

Po szkoleniach przeszliśmy do trzech intensywnych warsztatów. Pracowaliśmy metodami *design thinking* zaadaptowanymi do kontekstu polityki publicznej. Zaczynaliśmy od konkretnego, każdy zespół definiował problemy ze swojego biura, miejsca, w których generatywna AI może pomóc albo zaszkodzić. Z tych szczegółowych obserwacji stopniowo wyłanialiśmy ogólniejsze zasady. Karteczki na tablicy, dyskusje w małych grupach, synteza, znowu karteczki. Siedem zasad przewodnich, które znalazły się w finalnym dokumencie, nie zostało napisanych przez ekspertów i narzuconych z góry. Wyrosły z doświadczeń ludzi, którzy na co dzień pracują w urzędzie.

Z tych wszystkich dyskusji zmaterializował się prototyp dokumentu. Nie był on idealny, jego format, grafika, sporo treści – wszystko to miało się jeszcze zmienić. Do tego momentu trapiła mnie wątpliwość: czy z pracy kilkudziesięciu osób o różnych perspektywach może powstać spójny tekst? Okazało się, że może.

Feedback szedł falami

Później zaczął się najtrudniejszy etap – zbieranie opinii. Najpierw grupa robocza recenzowała własną pracę. Potem przyszła pora na ekspertów zewnętrznych, którzy nie uczestniczyli w warsztatach i patrzyli świeżym okiem. Ich uwagi były szczegółowe, akademicko rygorystyczne i bardzo przydatne.

Prawdziwy test przyszedł jednak od zupełnie innej grupy – urzędników, którzy nie brali udziału w tworzeniu dokumentu, a mieli go potem stosować w codziennej pracy. Przeprowadziłem z nimi wywiady. Jeden moment zapamiętam na długo: rozmówca zapytał wprost, czy dokument będzie zawierał „klauzulę sumienia” pozwalającą odmówić korzystania z generatywnej AI z powodów etycznych. To pytanie uświadomiło mi dwie rzeczy. Po pierwsze, że obawy związane z AI sięgają głębiej, niż zakładaliśmy. Po drugie, że język naszego dokumentu, wzbogacony przez eksperckie konsultacje, stał się momentami nazbyt akademicki dla przeciętnego pracownika. Informacja zwrotna od ludzi „z frontu” przypomniła nam, kto na co dzień będzie korzystał z opracowywanego przez nas dokumentu i jak powinniśmy go dostosować.

„Kierunki GenAI” to obszerny dokument ze szczegółowymi wytycznymi, siedmioma zasadami przewodnimi i szerokim kontekstem merytorycznym – od wyjaśnienia, czym jest generatywna AI, po konkretne scenariusze użycia i ryzyka. Taki dokument jest jednak zbyt rozbudowany, żeby mieszkańcy mogli szybko sprawdzić, jak ich miasto podchodzi do generatywnej sztucznej inteligencji. Dlatego obok „Kierunków GenAI” powstał pięciostronicowy „Warszawski Kodeks GenAI”, w którym siedem zasad zostało skondensowanych do czterech, ujętych tak, żeby każdy mógł je przeczytać w dziesięć minut i wiedzieć, o co chodzi. To właśnie Kodeks, jako wizytówkę podejścia Warszawy, przetestowaliśmy z mieszkańcami miasta. Pytaliśmy o zrozumiałość i wiarygodność: czy te zasady są dla was jasne? Czy ufacie, że miasto podchodzi do tematu poważnie?

Oba dokumenty zaprezentowaliśmy 27 stycznia 2025 r. podczas konferencji w Centralnym Domu Technologicznym. Wydarzenie było transmitowane na żywo i tłumaczone na język migowy. Fakt, że sala wypełniła się po brzegi, był dla nas jasnym sygnałem: temat sztucznej inteligencji przestał być niszowy.

Według naszej wiedzy był to pierwszy w Europie kompleksowy dokument samorządowy tego typu, dlatego od początku przygotowaliśmy wersję anglojęzyczną, którą oficjalnie ogłosiliśmy na konferencji MIT Global Startup Workshop 2025. Zainteresowanie dokumentem wyraziły nie tylko Gdynia, Poznań, Rzeszów czy Zduńska Wola, lecz

również zagraniczni partnerzy, jak Wiedeń, Wilno czy Monachium, którzy zaproponowali wymianę doświadczeń.

Po premierze grupa robocza przekształciła się w stałe ciało eksperckie z podgrupami tematycznymi, a do jej pierwotnego składu dołączyły nowe jednostki miejskie. W ramach tej szerszej grupy, dzięki wspólnej wymianie wiedzy i pracy nad konkretnymi wyzwaniami, zaczęły powstawać pierwsze projekty przełożenia zasad na praktykę. Potwierdziło to tezę leżącą u podstaw całego projektu: ludzie, którzy współtworzą zasady, potrafią je potem stosować.

Co wynoszę z tego procesu po ponad roku pracy? Przede wszystkim przekonanie, że ustalenie zasad przed wdrożeniem nie spowalnia, tylko przyspiesza prace. Kiedy ludzie wiedzą, co wolno i gdzie przebiegają granice, nie muszą się tego domyślać przy każdej czynności. Nie piszą maili z pytaniem, czy mogą użyć Claude AI do przygotowania notatki. Mają ramy, więc działają.

” *Przed rozpoczęciem projektu obawiałem się, że podejście „najpierw etyka” będzie odebrane jako hamulec. Okazało się odwrotnie: jasne zasady budują zaufanie, zarówno wśród pracowników, jak i na zewnątrz organizacji.*

Nauczyłem się też, że współtworzenie jest czasochłonne. Zebranie dziewiętnastu osób z ośmiu biur, zorganizowanie szkoleń, warsztatów, rund konsultacji to miesiące pracy, które można by skrócić, gdyby jedno biuro napisało wytyczne i rozesłało gotowe instrukcje. Tyle że ludzie, którzy sami kształtowali dokument, traktują go jako swój. Nie potrzebują zarządzenia, żeby go stosować, i to jest coś, czego ogólnym komunikatem się nie osiągnie.

Sam dokument to zresztą tylko część efektu. Drugą jest sieć ludzi, specjalistów z różnych biur, którzy rozumieją temat, potrafią o nim rozmawiać i wiedzą, do kogo zadzwonić, gdy pojawi się problem. Ta sieć okazała się równie wartościowa co sam tekst.

„Kierunki GenAI” są dokumentem żywym i będą się zmieniać wraz z technologią, regulacjami i potrzebami miasta, ale podejście, metoda i sposób myślenia zostają. Każdy samorząd może to powtórzyć niezależnie od wielkości i zasobów. Bazą może być nasz warszawski dokument, jednak najważniejsze jest powiązać ludzi, posadzić ich przy jednym stole, dać im czas i zacząć od pytania, jakie wartości chcemy chronić.