

Część III – Miary, oceny i wskaźniki funkcjonowania chatbotów

Wprowadzenie do Części III

W następstwie pomyślnie przeprowadzonej implementacji chatbot działa samodzielnie. Początkowo bardzo intensywnie jest nadzorowany przez tzw. botmastera, który kontroluje poprawność jego działania i koryguje pojawiające się błędy. Botmaster ma wgląd w wewnętrzne procesy chatbota: może przeglądać prowadzone przez niego rozmowy oraz statystyki i wykresy dotyczące tych rozmów. Chatbot wyposażony jest standardowo w mechanizm raportowania, który zbiera szczegółowe dane o przeprowadzonych rozmowach i wykonanych zadaniach. W przypadku ewentualnych, pomyłek w połączeniach skryptów, ścieżek dostępu do plików, jakichkolwiek nieprawidłowości realizacji implementacji, botmaster poprawia wewnętrzną konfigurację bazy wiedzy oraz odpowiednio rozwija bazę wiedzy chatbota. Wraz z upływem czasu nadzór botmastera przechodzi w regularną konserwację systemu i okresowe testy, dużo mniej intensywne, aniżeli tuż po implementacji.

Jakimi cechami powinien się charakteryzować zaimplementowany chatbot? Jak zdiagnozować występowanie technicznych problemów? W którym momencie określić potrzebę zmodernizowania technologii? Z jednej strony istnieją uzasadnione przyczyny dla modelowania i odzwierciedlenia ludzkiego zachowania w działaniu chatbota, reprezentującego na stronie internetowej żywą, myślącą istotę. Z drugiej strony chatbot wciąż pozostaje narzędziem informatycznym, więc charakterystyka jego projektu powinna podlegać założeniom i specyfikacjom systemu informatycznego wykoncypowanego w fazie przygotowawczej. Posiadając niezależność i autonomię pojedynczej wirtualnej osoby, będąc równocześnie narzędziem informatycznym ułatwiającym komunikację między firmą a użytkownikiem, chatbot stał się elementem promowania innowacyjnej komunikacji w systemach interakcji człowiek-komputer, odmiennej od tej wyłącznie międzyludzkiej.⁵⁷

Newralgicznym punktem każdej implementacji chatbota jest opracowywanie jego bazy wiedzy. W zależności od użytego modułu przetwarzania języka naturalnego oraz dziedzin, w których chatbot ma się poruszać, istotne jest odpowiednie dobranie i zróżnicowanie wypowiedzi. Jest to proces najbardziej czasochłonny ze wszystkich etapów implementacji. W

⁵⁷ Gratch J., Marsella S., Egges A., Eliens A., Isbister K., Paiva A., Rist T., ten Hagen P., *Design criteria, techniques and case studies for creating and evaluating interactive experiences with virtual humans*, AAMAS Workshop on Embodied Conversational Agents: Balanced Perception and Action, New York, 2004, s. 1-2.

przypadku błędnego oszacowania (najczęściej niedoszacowania) ilości czasu potrzebnego na zebranie i wprowadzenie wszystkich danych, może nastąpić przedłużenie czasu realizacji implementacji i przesunięcie terminu realizacji całości projektu. Podobnym problemem może być konieczność modyfikacji pierwotnego projektu w przypadku nieplanowanego zaistnienia różnych okoliczności zewnętrznych, np. zmiana nazwy, danych teleadresowych firmy itp.

Implementacja chatbota powinna cechować się odpowiednią jakością funkcjonowania, żeby móc sprostać oczekiwaniom użytkowników, będących klientami danej firmy. Problem błędnego zrozumienia wypowiedzi użytkownika jest bardzo trudny do przezwyciężenia. Nieprawidłową interpretację zapytania użytkownika powinien wykryć i naprawić odpowiednio do tego skonfigurowany mechanizm zarządzania dialogiem. Z kolei związane z tym problemem niezdecydowanie chatbota wyrażane w ponawianych prośbach o zredefiniowanie wypowiedzi również powinno podlegać monitorowaniu i naprawieniu, gdyż w innym przypadku wywołuje to frustrację użytkownika. Przy częstym występowaniu takich sytuacji, dobrze żeby chatbot zaprezentował kilka możliwych interpretacji wypowiedzi, z których użytkownik mógłby wybrać odpowiednią dla siebie.

Implementację chatbota można ocenić pod względem technicznym i technologicznym biorąc pod uwagę: informacje przez niego przekazywane oraz realizowane zadania, znaczenie gestów, mimiki i wizualizacji, zrozumiałe dla użytkownika zachowanie oraz naturalność osadzenia chatbota w danym środowisku systemowym. Dodatkowo jeszcze można dodać szerszej zakrojone kryteria, mianowicie: estetyka projektu chatbota powinna być zorientowana wokół jego zasadniczych funkcji, rozwój chatbota powinien być wspierany przez najnowsze technologie multimedialne. Chatbot jest naturalnie i przekonująco odbierany przez użytkowników najczęściej wówczas, gdy odpowiednia graficzna jakość połączona z kontekstualną zawartością dobrze komponują się ze środowiskiem implementacji.⁵⁸ Ewentualne problemy techniczne, mogące się pojawić podczas realizacji implementacji, są takie same jak w przypadku implementacji każdego innego systemu informatycznego, związane z zapewnieniem właściwej funkcjonalności, stabilności działania systemu oraz zapewnienia bezpieczeństwa dla zbieranych i przetwarzanych informacji.

⁵⁸ Eliens A., Huang Z., Hoorn J. F., Visser C., *ECA Perspectives – Requirements, applications, technology*, Dagstuhl Seminar Proceedings: Evaluating Embodied Conversation Agents, 2004, s. 1-2.

Zazwyczaj z pojęciem oceny implementacji systemu informatycznego wiąże się termin efektywności tego przedsięwzięcia. Określa ona relację między uzyskanymi wynikami a nakładami poniesionymi na ich osiągnięcie. W literaturze można znaleźć wiele metod oceny efektywności przedsięwzięć informatycznych. Należą do nich różnorodne metody: ROI, PP, NPV, IRR, PI, TCO, BSC, które można podzielić na proste, dyskontowe i wielokryterialne. Jednakże termin „efektywność” jest niejednoznaczny, często odmiennie rozumiany w literaturze przedmiotu, na przykład jako „sprawność”, „wydajność” czy „skuteczność”.⁵⁹ Z kolei informacje w branżowej prasie dotyczące implementacji poszczególnych chatbotów są najczęściej tematem mikro-ewaluacji⁶⁰, testującej zachowanie i oceniającej szeroko pojętą użyteczność danego chatbota. Zatem kwestia „przydatności” albo „użyteczności” chatbota okazuje się być zbyt ogólnie postawionym pytaniem.

Analiza i ocena chatbota nie może być oderwana od roli, jaką ma pełnić w danym środowisku i wykonywanych zadań, a także należy wziąć pod uwagę cechy charakterystyczne docelowych rozmówców.⁶¹ Jeden chatbot będzie głównie redukował ponoszone przez firmę koszty rozmów telefonicznych do klientów oraz liczbę obsługiwanych wiadomości e-mail, na które pracownicy działu obsługi klienta muszą nieustannie odpowiadać, z kolei inny będzie miał za zadanie identyfikację najczęściej pojawiających się tematów rozmów zaczynanych przez klientów i ich zapytań związanych z działalnością swego pracodawcy, np. firmy ubezpieczeniowej lub wydawniczej. Widać więc wyraźnie, jak trudne jest porównywanie względem siebie chatbotów - każdy z nich ma swój własny, niepowtarzalny profil działalności. Podczas ewaluacji chatbota, jeśli tylko to możliwe, należy brać pod uwagę każdy potencjalny rodzaj danych, które mogą być pomocne w dalszych jego ocenach i porównaniach. Dopiero poszczególne oceny zebrane w całość mogą zostać w ogólniejszym kontekście prześledzone i stanowić narzędzie informatyczne dostarczające dedykowanych rozwiązań szeroko rozumianej obsłudze klienta.

Nakłady poniesione na zakup i wdrożenie narzędzia informatycznego dość łatwo poddają się wycenie. Koszty, które wówczas się pojawiają, można podzielić na jednorazowe i bieżące.

⁵⁹ Nowicki A. (red.), *Komputerowe wspomaganie biznesu*, Wydawnictwo PLACET, Warszawa 2006, s.210-214.

⁶⁰ Buisine S., Hartmann B., Mancini M., Pelachaud C., *Conception et Evaluation d'un Modele d'Expressivité pour les Gestes des Agents Conversationnels*, Revue en Intelligence Artificielle RIA, Special Edition “Interaction Emotionnelle”, nr 20(4-5)2006, s. 2.

⁶¹ Xiao J., Stasko J., Catrambone R., *Embodied conversational agents as a UI paradigm: A framework for evaluation*, Proceedings of the AAMAS02 Workshop on Embodied conversational agents – let’s specify and evaluate them!, Bologna 2002, s. 1.

Jednorazowy jest zakup odpowiedniego oprogramowania wraz z licencjami, uruchomienie oprogramowania oraz przeszkolenie personelu do jego obsługi. Koszty bieżące składają się natomiast na eksploatację narzędzia informatycznego, jego rozwój i konserwację. Opcjonalnie mogą wystąpić trudno mierzalne koszty związane z chaosem organizacyjnym i informacyjnym występującym w początkowym okresie funkcjonowania narzędzia informatycznego, tuż po jego implementacji.⁶²

Trudno porównywać i oceniać między sobą koszty poniesione na zakup chatbota. Niski koszt jednej implementacji może być po prostu spowodowany zapotrzebowaniem na mniejszą, niekoniecznie rozbudowaną bazę wiedzy tego konkretnego przypadku. Z kolei wysoki koszt innej implementacji może wynikać z bardziej zaawansowanego projektu zastosowania chatbota. Różnica w cenie zależy od stopnia zaawansowania chatbota oraz przewidywanych funkcjonalności, które ma realizować.

Niezwykle trudnym zadaniem jest także wyczerpująca ocena chatbota. Jest to narzędzie, dotyczące tak wielu dziedzin (informatyka, NLP, psychologia, kognitywistyka, UI/UX, grafika, synteza mowy), że trudno przedstawić i zgłębić wszystkie aspekty jego budowy, implementacji i funkcjonowania. Ujednolicone podejście badawcze utrudnia niestety fakt, że każda z firm wdrożeniowych pracuje z własną, unikalną technologią, opierając się na wypracowanej przez siebie autorskiej metodologii realizacji implementacji chatbota.

Chatboty są niezwykle zróżnicowane. Implementowane są w kanałach komunikacyjnych i w procesach wewnętrznych firm różnych dziedzin: nieruchomości, ubezpieczenia, kosmetyki, media, edukacja itd. Różnią się też między sobą zarówno cechami konstrukcji wewnętrznej: objętość bazy wiedzy, moduł NLP, interakcja ze środowiskiem oraz cechami zewnętrznymi: sposób osadzenia na stronie, technologia użyta do wizualizacji itp. Zasadne staje się więc pytanie, czy można w ogóle myśleć o stworzeniu uniwersalnego podejścia do oceny implementacji chatbota?

⁶² Nowicki A. (red.), *Informatyka dla ekonomistów. Studium teoretyczne i praktyczne*, PWN, Warszawa-Wrocław 1998, s. 75.

Rozdział 5. Miary oceny wdrożenia systemów informatycznych niekonwersacyjnych oraz opartych o model konwersacyjny

Ostatnim etapem tworzenia systemu informatycznego jest jego ewaluacja. W literaturze przedmiotu ocena przedsięwzięć informatycznych jest określana przez charakterystyki takie jak: czas, sposób i przedmiot oceny, jakość oprogramowania, niezawodność, zgodność z wymaganiami czy bezpieczeństwo. Charakterystyki te w stosunkowo ogólny sposób oceniają przedsięwzięcia informatyczne. Uważa się, że projekt został zakończony sukcesem, jeśli jego wykonanie nastąpiło zgodnie z harmonogramem, budżetem i przewidzianą specyfikacją projektu.⁶³ Jednak wymienione miary nie wyczerpują tematu ewaluacji przedsięwzięcia informatycznego i nie są wystarczające do weryfikacji sukcesu danego projektu. Istnieje wiele metod służących do dokładniejszej oceny jakości wdrożonego i funkcjonującego systemu.

Efektywność systemu jest jedynym z elementów jego oceny. Efektywność rozumiana jest jako stosunek uzyskanych korzyści do poniesionych kosztów. KPI (Key Performance Indicators) to kluczowe wskaźniki efektywności. Do KPI zaliczają się finansowe i niefinansowe wskaźniki służące do pomiaru realizacji celów organizacji. Firmy mogą za ich pomocą mierzyć i sprawdzać, w jaki sposób wdrożenie systemu informatycznego wpływa na osiąganie celów biznesowych, na przykład poprzez monitorowanie czasu odpowiedzi systemu.⁶⁴ Wskaźniki KPI dobierane są indywidualnie, w zależności od potrzeb firmy, branży i jej cech specyficznych. Jednak nie ma konkretnego zestawu takich wskaźników dla firmy, co może uzasadniać wiele istniejących miar oceny projektów, w tym systemów informatycznych.

Za jedną z najbardziej znanych i najczęściej opisywanych miar w literaturze uznaje się ISO/IEC 25010:2011.⁶⁵ Ten model oceny ma zastosowanie zarówno do systemów komputerowych, jak i oprogramowania.⁶⁶ Jest to zaktualizowana wersja oryginalnej normy

⁶³ Lech P. *Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją*. Gdańsk 2007. s.72.

⁶⁴ Gryczuk A. Kluczowe wskaźniki efektywności (KPI) jako narzędzie doskonalenia efektywności operacyjnej film produkcyjnych zorientowanych na lean. 2011. s.28-31.

⁶⁵ Polillo R. *Quality Models for Web [2.0] Sites: a Methodological Approach and a Proposal*. Mediolan 2011.

⁶⁶ <https://www.iso.org/standard/35733.html> [dostęp 08.2022]

ISO/IEC 9126, która rozszerzyła model bazowy o dwie nowe charakterystyki: bezpieczeństwo i zgodność systemu. Została utworzona przez Międzynarodową Organizację Normalizacyjną w 1991 roku i jest to międzynarodowa norma, będąca oceną jakości oprogramowania. Norma ta składa się z czterech części⁶⁷:

1. Model jakości – model klasyfikuje jakość oprogramowanie używając cech i ich podkategorii.
2. Miary zewnętrzne – miary skupiające się na uruchomionym, działającym oprogramowaniu.
3. Miary wewnętrzne – miary statyczne, nieskupiające się na wykonaniu oprogramowania.
4. Miary jakości użytkowej – miary stosowane, gdy oprogramowanie używane jest jako produkt końcowy w rzeczywistych warunkach.

Początkowa wersja ISO/EIC 9126 w ogólny sposób oceniała jakość systemu, opierała się na sześciu głównych cechach jakościowych, które zawierają dwadzieścia siedem charakterystyk:

1. Funkcjonalność – oprogramowanie posiada funkcje zgodne z wymaganiami użytkownika.
2. Niezawodność – oprogramowanie umożliwia wykonywanie założonych zadań w ustalonych warunkach.
3. Użyteczność – oprogramowanie jest łatwe do zrozumienia, nauki i jest łatwe w użyciu.
4. Wydajność – oprogramowanie jest wydajne zgodnie z wcześniejszymi założeniami.
5. Możliwość konserwacji – w oprogramowaniu dokonuje się zmian, poprawianie błędów, polepszaniu i dostosowywaniu.
6. Przenośność – oprogramowanie można przekształcić do użycia w innym środowisku systemowym.

W aktualnej wersji ISO/IEC 25010:2011 na model jakości systemu składają się miary wewnętrzne i zewnętrzne systemu. Obejmuje ona osiem cech głównych i trzydzieści jeden charakterystyk. Tak jak zilustrowano to w Tabeli 4., dodane zostały dwie cechy. Uznano, że nową cechą zostanie bezpieczeństwo oprogramowania i będzie mieć na nią wpływ poufność, integralność, niezaprzeczalność, odpowiedzialność oraz autentyczność systemu. Pomimo iż

⁶⁷ Botella P., Burgués X., Carvallo J.P., Franch X. ISO/IEC 9126 in practice: what do we need to know?. 2004, s.3.

zgodność jest również nową cechą, to jej charakterystyki zostały zaczerpnięte z innych cech, a mianowicie: współlistnienie zaczerpnięto z przenośności, a współdziałanie zaczerpnięto z funkcjonalności.⁶⁸

Tabela 4. Porównanie charakterystyk pomiędzy ISO/EIC 9126 i ISO/IEC 25010:2011.

ISO/EIC 9126	ISO/IEC 25010:2011
Funkcjonalność (Functionality)	Przydatność funkcjonalna (Functional suitability)
Niezawodność (Reliability)	Niezawodność (Reliability)
Użyteczność (Usability)	Użyteczność (Usability)
Wydajność (Efficiency)	Wydajność (Performance efficiency)
Możliwość konserwacji (Maintainability)	Możliwość konserwacji (Maintainability)
Przenośność (Portability)	Przenośność (Portability)
-	Bezpieczeństwo (Security)
-	Zgodność (Compatibility)

Źródło: Opracowanie własne na podstawie: Polillo R. Quality Models for Web [2.0] Sites: a Methodological Approach and a Proposal. Mediolan 2011.

Przytoczona miara ISO/EIC skupia się na badaniu ogólnej jakości systemów informatycznych. Natomiast ocena efektywności przedsięwzięcia informatycznego dotyczy zazwyczaj kilku poziomów.⁶⁹ Klasyfikacja korzyści i kosztów jest najważniejszą analizą przedsięwzięć informatycznych uwzględniającą ekonomiczne wyniki przedsiębiorstwa.⁷⁰ W zamieszczonej dalej Tabeli 5. przedstawiona została klasyfikacja korzyści i kosztów zastosowania rozwiązań informatycznych w przedsiębiorstwie.

Mając na uwadze związek zastosowań informatyki i ekonomicznych wyników przedsiębiorstwa, korzyści i koszty dzielone są na bezpośrednie i pośrednie. Bezpośrednie korzyści (analogicznie w przypadku kosztów) powstają wtedy, gdy dzięki użytej technologii informacyjnej w przedsiębiorstwie nastąpiło polepszenie wyników ekonomicznych. Natomiast o korzyściach pośrednich mówi się, gdy wpłynęły korzystnie na działanie

⁶⁸ Polillo R. Quality Models for Web [2.0] Sites: a Methodological Approach and a Proposal. Mediolan 2011.s.3.

⁶⁹ Szplit M., Ekonomiczne aspekty działania systemowa informatycznych w przedsiębiorstwach, Kielce 2012, s.1.

⁷⁰ Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007.s.38.

przedsiębiorstwa, ale nie przyczyniły się bezpośrednio do polepszenia jej wyników ekonomicznych.

Kolejnym kryterium podziału jest możliwość pomiaru kosztów i korzyści, mogą być one mierzalne lub niemierzalne. Korzyści bądź koszty są mierzalne, wtedy, gdy można je zapisać miarą finansową bądź fizyczną. Problem doboru miary wyników niemierzalnym rozwiązuje się m.in. określając wyniki miarami ilościowymi i jakościowymi. Koszty i korzyści mogą zostać ujęte finansowo i niefinansowo. Koszty bądź korzyści finansowe można także ująć w jednostkach pieniężnych.

Tabela 5. Klasyfikacja korzyści i kosztów zastosowania rozwiązań informatycznych w przedsiębiorstwie.

Kryterium klasyfikacji	Możliwe wartości
Związek z wynikami ekonomicznymi organizacji	• Bezpośrednie • Pośrednie
Możliwość pomiaru	• Mierzalne • Niemierzalne
Możliwość przypisania wartości pieniężnej	• Finansowe • Niefinansowe

Źródło: Opracowanie własne na podstawie: Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007.

Zgodnie z powyższą klasyfikacją kosztów i korzyści, ocena efektywności przedsięwzięcia informatycznego może występować na kilku poziomach, w zależności od uwzględnionych kryteriów w analizie. Efektywność ze względu na kryterium między przedsięwzięciem a wynikami ekonomicznymi organizacji można mierzyć w następujący sposób:⁷¹

1. Na poziomie bezpośrednich korzyści i kosztów.
2. Na poziomie pełnych korzyści i kosztów (bezpośrednie i pośrednie).

Ze względu na efekty podjęcia od strony finansowej efektywność ocenia się:

1. Finansowe korzyści i koszty.
2. Pełne korzyści i koszty (finansowe i niefinansowe).

⁷¹ Ibidem, s.179.

Łącząc dwie powyższe klasyfikacje powstaje kombinacja ocen efektywności na poziomach przedstawionych poniżej w zamieszczonej dalej Tabeli 6.

Tabela 6. Sposób pomiaru efektywności przedsięwzięć informatycznych.

Poziom pomiaru	Klasyfikacja pomiarów	Sposób wyrażenia pomiaru efektywności
Mierzalne koszty i korzyści	Bezpośrednie, finansowe	Jednostki pieniężne
	Bezpośrednie, pełne (finansowe i niefinansowe)	Jednostki pieniężne: koszty niefinansowe wyrażone w miarach wartościowych, rangi punktowe
	Pełne (bezpośrednie i pośrednie), finansowe	Jednostki pieniężne
	Pełne (bezpośrednie pośrednie, finansowe, niefinansowe)	Rangi punktowe
Niemierzalne koszty i korzyści	Niemierzalne koszty i korzyści	Opis słowny

Źródło: Opracowanie własne na podstawie: Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007.

Do najłatwiej mierzalnych korzyści i kosztów zalicza się bezpośrednie i finansowe korzyści i koszty. Można je wyrazić w wymiarze pieniężnym oraz wynikają one bezpośrednio z wdrożenia rozwiązania informatycznego. Takie korzyści najczęściej powstają dzięki automatyzacji prac w przedsiębiorstwie bądź skróceniu czasu trwania procesów gospodarczych, natomiast koszty mogą dotyczyć budowy infrastruktury informatycznej czy opłacenia licencji.

Drugą możliwą kombinacją ocen efektywności są bezpośrednie oraz finansowe i niefinansowe korzyści i koszty. Pomimo, iż jest to możliwy scenariusz, najczęściej bezpośredni wpływ przedsięwzięcia informatycznego jest wyrażany w mierzalnych finansowo jednostkach i modelach. Jednak, jeżeli nie jest możliwe wyrażenie korzyści i kosztów do postaci wartościowej, stosuje się rangi punktowe.

Trzecią możliwością jest wystąpienie pełnych (bezpośrednich i pośrednich) kosztów oraz korzyści finansowych. Korzyści pośrednie mogą wynikać z lepszej informacji, z produkowania nowych usług i zapewnienia lepszej jakości obsługi klientów. Do kosztów pośrednich zalicza się czas pracowników w ujęciu wartościowym, czyli dodatkowy czas jaki pracownicy poświęcili na pomoc lub naprawę błędów, a także zmniejszone przychody spowodowane przez same błędy w obsłudze systemu. Czwartą możliwą opcją pomiaru efektywności przedsięwzięć informatycznych jest poziom pełnych kosztów i korzyści. Jedyną możliwą opcją ich pomiaru jest użycie rang punktowych.

Ekonomiczne aspekty oceny przedsięwzięć informatycznych przedstawione w literaturze przedmiotu skupiają się na kryteriach metod i narzędziach pomiaru kosztów i korzyści. Lech w swoim artykule połączył dwie klasyfikacje ekonomicznego podziału kosztów: stworzoną przez Renkema, Berghouta oraz Mayora^{72,73}. Na ich podstawie zaproponował podział metod oceny przedsięwzięć informatycznych biorąc pod uwagę sposób pomiaru ich korzyści i kosztów. Istnieje wiele kryteriów klasyfikacji podziału tych metod oceny, które przedstawiono w Tabeli 7. Widać w niej następujące klasyfikacje: sposób ujęcia kosztów i korzyści, poziom obserwacji, zakres oceny i specyfikę technologii informatycznej.

Tabela 7. Metody oceny przedsięwzięć informatycznych na podstawie pomiaru korzyści i kosztów.

Kryterium klasyfikacji	Możliwe wartości
Sposób ujęcia kosztów i korzyści, technika pomiaru	Metody ilościowe: 1. Deterministyczne (finansowe) 2. Probabilistyczne Metody jakościowe: 1. Multikryterialne 2. Analizy strategicznej
Poziom obserwacji	1. Pojedynczy projekt 2. Porównanie projektów 3. Cała organizacja

⁷² Renkema T., Berghout E. (1997) Methodologies for information systems investment evaluation at the proposal stage: a comparative review, Information and Software Technology 39/1997. Eindhoven 1997. s. 1-13.

⁷³ Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007.s.72.

Zakres oceny	1. Wszystkie aspekty 2. Wybrane aspekty
Specyfika TI	1. Metoda specyficzna dla TI 2. Metoda zaadaptowana z innych zastosowań

Źródło: Opracowanie własne na podstawie: Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007.

W klasyfikacji obejmującej sposoby ujęcia kosztów i korzyści widnieje podział na metody ilościowe i jakościowe. Na metody ilościowe składają się metody deterministyczne, nazywane również finansowymi, które odzwierciedlają korzyści i koszty poprzez pojedyncze wartości. Metody probabilistyczne wykorzystują narzędzia statystyki matematycznej i rachunku prawdopodobieństwa, korzyści i koszty traktowane są jako zmienne losowe. Druga grupa metod, jakimi są metody jakościowe, w ocenie uwzględniają również czynniki niefinansowe. Do metod jakościowych zalicza się metody multikryterialne, które uwzględniają finansowy i niefinansowy element oceny oraz metody analizy strategicznej.

W następnych akapitach zostaną przeanalizowane wybrane ekonomiczne metody ilościowe i jakościowe przedstawione w zamieszczonej dalej Tabeli 8.

Tabela 8. Metody ilościowe i jakościowe w ocenie przedsięwzięć informatycznych na podstawie pomiaru korzyści i kosztów.

Metody ilościowe	Metody jakościowe
1. NPV 2. ROI 3. TCO 4. TEI Inne: EVA, PP	1. Ekonomia Informacji 2. Zrównoważona Karta Wyników Inne: Metoda Pięciu Filarów, Analiza portfelowa

Źródło: Opracowanie własne na podstawie: Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007.

NPV – NET PRESENT VALUE

NPV, czyli bieżąca wartość netto to miara dyskontowanych przepływów pieniężnych netto. Bieżąca wartość netto to metoda oceny efektywności ekonomicznej inwestycji. NPV jest

interpretowana na podstawie wyliczonego wskaźnika. Wskaźnik obliczany jest poprzez różnicę przepływów gotówkowych (z uwzględnieniem wartości pieniądza w czasie) i nakładów początkowych. Bieżąca wartość netto umożliwia na bezwzględną ocenę projektu. Projekt jest opłacalny, gdy wartość wskaźnika jest większa od zera. Minusem prostoty tej metody jest stosowanie podczas trwania projektu stałej stopy dyskontowej i uproszczenie dotyczące ryzyka projektu.⁷⁴

ROI – RETURN OF INVESTEMENT

Bardziej złożoną analizowaną miarą jest zwrot z inwestycji (ang. Return of Investment, ROI). Jest to wskaźnik, który również jest używany do szacowania opłacalności projektów informatycznych.⁷⁵ ROI mierzy korzyści uzyskane po inwestycji w dany projekt informatyczny. Analiza zawiera materialne i niematerialne korzyści uzyskane z implementacji systemu do przedsiębiorstwa. Na zmianę wartości zwrotu z inwestycji wpływa: zmiana przychodów, zmiana kosztów, zmiana wydajności, korzyści niemierzalne, koszty nabycia systemu informatycznego. Przedsiębiorstwa prowadzą dokumentację dotyczącą wyżej wymienionych zmian, dlatego analiza ROI jest łatwo mierzalna. Jednak wadą wskaźnika zwrotu z inwestycji jest problem z identyfikacją, czy zmiany w przychodach związane były tylko z daną inwestycją, a nie wynikały z innych projektów. Dlatego do pomiaru wskaźnika ROI firmy powinny wybrać tylko te przychody, które są powiązane z danym projektem, na przykład wdrożeniem systemu informatycznego. Do wyliczenia wartości wskaźnika zwrotu z inwestycji należy podzielić uzyskane przychody przez uzyskane koszty.

TCO – TOTAL COST OF OWNERSHIP

Kolejną stosowaną miarą oceny systemu informatycznego jest TCO (Total Cost of Ownership), czyli metoda całkowitych kosztów posiadania. Została stworzona przez Gartner Group, amerykańskie przedsiębiorstwo analityczno-badawcze, które w głównej mierze zajmuje się strategicznym wykorzystywaniem technologii. Metoda całkowitych kosztów posiadania pozwala w dokładny sposób ująć koszty powstałe w wyniku posiadania rozwiązania informatycznego.^{76,77} Ta metoda w sposób całościowy skupia się na wszystkich

⁷⁴ Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007.s.76-77.

⁷⁵ Osuszek Ł. Ocena efektywności inwestycji w rozwiązania BPM. IBM 2011, s. 358.

⁷⁶ Mastalerz M. Podstawy szacowania kosztów wdrażania portali społecznościowych w przedsiębiorstwie. Szczecin 2018, s.82.

kosztach wytworzonych w procesie pozyskiwania nowych technologii informatycznych. TCO hierarchizuje koszty nabycia, posiadania oraz użytkowania systemu. Na wycenę TCO składa się wiele kosztów. Oprócz kosztów samego zakupu produktu, występują również koszty powstałe przy zakupie, badaniu kwalifikacji dostawcy, transportu, przechowywania czy utylizacji. Na sumę całkowitych kosztów składa się:⁷⁸

1. Kapitał trwały: sprzęt komputerowy, sieciowy, oprogramowanie
2. Koszty powstałe poprzez stratę czasu pracowników
3. Koszty powstałe przy tworzeniu oprogramowania
4. Koszty związane ze wsparciem technicznym
5. Koszty związane z nauką potrzebną do użytkowania: kursy obsługi, certyfikaty
6. Koszty niesformalizowanej nauki: książki
7. Koszty administracji
8. Koszty powstałe przy zarządzaniu danymi
9. Koszty dostaw i zapasów

Głównym celem w TCO jest znalezienie ukrytych kosztów zakupu produktu. Uznaje się, że cena zakupionego produktu to jedynie 25% całkowitych kosztów poniesionych podczas jego całego cyklu życia w firmie.⁷⁹ Do ukrytych kosztów zalicza się: szkolenie pracowników, koszty pracowników, koszt instalacji, preferencje użytkownika, nieprzewidziany przestój, zarządzanie ryzykiem, koszty gwarancji, materiały eksploatacyjne i części, koszt utrzymania, koszt dostawy.

TEI – TOTAL ECONOMIC IMPACT

Metoda rozszerzająca metodę całkowitych kosztów posiadania (TCO) oraz wskaźnik zwrotu z inwestycji (ROI) to metoda całkowitego wpływu ekonomicznego. Została opracowana przez firmę Forrester, jest szerzej znana pod angielską nazwą Total Economic Impact (TEI). TEI oprócz analizy kosztów i korzyści dodatkowo zajmuje się analizą ryzyka i elastyczności. Nowy czynnik, jakim jest ryzyko, ujmuje prawdopodobieństwo zdarzenia, że firma poniesie większe koszty niż planowała oraz korzyści wynikające z inwestycji będą niższe. Drugi nowy czynnik, jakim jest elastyczność uwzględnia wartości przyszłych możliwości i korzyści, które

⁷⁷ Lewicki W., (2017), Wskaźnik Total Cost of Ownership jako narzędzie służące optymalizacji procesu zakupowego samochodów osobowych. *Folia Pomeranae Universitatis Technologiae Stetinensis Oeconomica*

⁷⁸ Osuszek Ł. Ocena efektywności inwestycji w rozwiązania BPM. IBM 2011, s.357,358.

⁷⁹ Barbusova M., Gaso M., Medvecká I. Use of TCO Analysis in Industry 4.0. Pilzno 2019. s.12.

dana inwestycja może zapewnić firmie w przyszłości. Uwzględnienie tych czynników pomaga przedsiębiorstwu uzyskać bardziej realistyczny ogląd sytuacji, na który może się przygotować poprzez dokładniej dobraną strategię.⁸⁰

Ekonomika Informacji

Ekonomika Informacji jest miarą jakościową i uważana jest za pierwszą ekonomiczną metodę oceny przedsięwzięcia informatycznego, która powstała w 1988 roku. Oprócz tradycyjnych miar finansowych, ocenia wartość projektu w domenie biznesowej i technologicznej. Do oceny względnej wartości korzysta z rozszerzonej stopy zwrotu. Wartość projektu jest mierzona w skali 0 – 100, gdzie „100” oznacza projekt o największym znaczeniu dla organizacji. Elementy składające się na wartość rozwiązania informatycznego mają przyporządkowane rangi w zakresie 0 – 5. Następnie rangi mnożone są przez wagę, jaką mają dane elementy w procesie spełnienia celów organizacji. Końcowa wartość projektu obliczana jest w następujących pięciu krokach scharakteryzowanych poniżej⁸¹.

1. Ustalenie rozszerzonej stopy zwrotu i przypisanie jej rangi

Krok ten polega na obliczeniu tradycyjnej stopy zwrotu i stopniowym rozszerzaniu jej o elementy, związane z danym rozwiązaniem informatycznym. Do obliczenia tradycyjnej stopy zwrotu bierze się pod uwagę bezpośrednie korzyści finansowe, powstałe w wyniku wdrożenia projektu w firmie: uzyskane z projektu przychody, zaoszczędzone koszty, poniesione nakłady finansowe, koszty eksploatacji. Wyliczona stopa zwrotu zostaje poprawiana i może uwzględnić dodatkowe korzyści finansowe uzyskane podczas wdrożenia. Do końcowej stopy zwrotu zespół dokonujący oceny projektu powinien nadać rangę odzwierciedlającą wymagania przedsiębiorstwa dotyczące zwrotu z inwestycji.

2. Ustalenie rang dla miar domeny biznesowej

Na tym etapie oceniane są niefinansowe korzyści i koszty domeny biznesowej. Można do nich zaliczyć przewagę konkurencyjną czy zgodność ze strategią. Do wybranych korzyści i kosztów przypisywane są rangi od zera do pięciu.

⁸⁰ <https://ecapitaladvisors.com/blog/forresters-total-economic-impact-methodology/> [dostęp 08.2022]

⁸¹ Lech P. Metodyka ekonomicznej oceny przedsięwzięć informatycznych wspomagających zarządzanie organizacją. Gdańsk 2007, s.104-112.

3. Ustalenie rangi dla miar domeny technologicznej

Analogicznie do kroku drugiego, na tym etapie ustala się rangi dla miar domeny technologicznej. Do tych miar należy między innymi ryzyko infrastrukturalne, zmienność początkowego celu projektu.

4. Ustalenie systemu wartości organizacyjnej – wag

Organizacja określa priorytety danego projektu informatycznego. Miarom nadawane są wagi tak, aby suma wag pomnożona przez maksymalną wartość rangi równą 5 równała się 100.

5. Ustalenie miary wartości projektu

Po nadaniu wag poszczególnym miarom, następuje etap wyliczenia wartości ocenianego projektu informatycznego. Do wyliczenia tej wartości należy pomnożyć rangi przez ustalone wagi. Końcowa wartość projektu to suma wszystkich poszczególnych wartości miar.

Przytoczona Ekonomia Informacji wykorzystuje finansowe, niefinansowe oraz bezpośrednie i pośrednie koszty i korzyści do oceny rozwiązania informatycznego. W swojej ocenie uwzględnia też ryzyko, które pojawia się podczas realizacji projektu. Jest to jednak podejście bardzo subiektywne i ocena całego projektu zależy od osób oceniających dane przedsięwzięcie. Jej kolejną wadą jest zbyt duża ogólnikowość, która może nie być w stanie scharakteryzować istotnych elementów. Została jednak stworzona 34 lata temu i powinna zostać dostosowana do obecnych realiów technologicznych, które od 1988 zmieniły się diametralnie.

BS – BALANCED SCORECARD

Literatura przedmiotu wymienia również metodę zrównoważonej karty wyników (Balanced scorecard - BS) jako ekonomiczną miarę oceny systemu informatycznego. Metoda zrównoważonej karty wyników to instrument finansowy, który usprawnia mierzenie wyników przez firmy. Instrument ten pomaga w określeniu efektywności działań firmy w realizowanej strategii. Zgodnie z założeniami metody BS, strategię przedstawia się przy pomocy mierzalnych celów, które są niezbędne do realizacji zamierzonego planu. Cele firmy zawarte są w czterech perspektywach, które są mierzone za pomocą wybranych elementów finansowych i pozafinansowych.⁸²

⁸² Kuchta D., Ryńca R. Zrównoważona karta wyników i zrównoważona karta działania. Badania operacyjne i decyzje. 2007, s.94.

Na metodę zrównoważonej karty wyników składa się:

1. Perspektywa finansowa
2. Perspektywa klienta
3. Procesów wewnętrznych
4. Infrastruktury i rozwoju

Perspektywa finansowa wyszczególnia kilka wybranych środków finansowych, które sprawdzają powodzenie projektu. Zalicza się do niej miary takie jak: zwrot z kapitału, zmiana wartości dla akcjonariuszu oraz wykorzystanie posiadanych aktywów. Perspektywa klienta bierze pod uwagę co jest ważne dla klientów i interesariuszy oraz w jaki sposób firma powinna wyglądać, aby osiągnąć postawione cele. Ta perspektywa zwraca uwagę na cechy produktów, usług, relacje z klientami oraz wizerunek i reputację przedsiębiorstwa. Zadaniem perspektywy procesów wewnętrznych jest odpowiedzenie na pytanie w jakich procesach biznesowych nasza firma ma się wyróżniać. W skład jej oceny wchodzi środki i miary takie jak: rozwijanie produktów oraz usług, dostarczanie produktów oraz usług i usługi posprzedażowe. Czwarta perspektywa, jaką jest perspektywa uczenia się i rozwoju, odpowiada za ciągły rozwój firmy i wprowadzanie innowacji. W tej płaszczyźnie zwraca się uwagę na możliwości pracowników oraz systemu informacyjnego, motywację oraz czas na wprowadzenie produktu na rynek.⁸³

Opisane miary oceny systemu informatycznego nie wyczerpują wszystkich miary występujące w literaturze. Jednak zostały wybrane na podstawie przeanalizowanej literatury przedmiotu jako te najczęściej pojawiające się i najszerzej zbadane metody. W zamieszczonej dalej Tabeli 9. każdą z miar: metodę TCO, wskaźnik ROI, metodę TEI oraz metodę BS scharakteryzowano najważniejszymi cechami.

Tabela 9. Wybrane miary oceny systemu informatycznego – podsumowanie.

Miara oceny systemu informatycznego	Najważniejsze cechy
Bieżąca wartość netto - NPV	Miara zdyskontowanych przepływów pieniężnych netto, interpretowana na podstawie wskaźnika obliczanego na podstawie różnicy przepływów gotówkowych i nakładów początkowych.

⁸³ Chavan M. The balanced scorecard: a new challenge, Journal of Management Development. 2009.

Zwrot z inwestycji - ROI	Zawiera materialne i niematerialne uzyskane korzyści. Wpływ na zmianę wartości ROI: zmiana przychodów, zmiana kosztów, zmiana wydajności, korzyści niemierzalne, koszty nabycia systemu informatycznego.
Metoda pełnych kosztów posiadania - TCO	Suma wszystkich kosztów posiadania, użytkowania i zakupu. W kosztach należy uwzględnić koszty powstałe przy zakupie, badania kwalifikacji dostawcy, transportu, przechowywania czy utylizacji.
Metoda całkowitego wpływu ekonomicznego - TEI	Ulepszona wersja TCO i ROI, ocenia wpływ inwestycji w technologię. Określa koszty korzyści, elastyczność, czynniki ryzyka wpływające na decyzję inwestycyjną.
Ekonomika Informacji	Oprócz tradycyjnych miar finansowych, ocenia wartość projektu w domenie biznesowej i technologicznej. Do oceny względnej wartości korzysta z rozszerzonej stopy zwrotu. Wartość projektu jest mierzona w skali 0 – 100, gdzie „100” oznacza projekt o największym znaczeniu dla organizacji.
Metoda zrównoważonej karty wyników - BS	Mierzy wynik firmy i efektywności. Definiuje ona cele i mierniki służące do analizy postępu projektów, które dzieli się na cztery perspektywy: finansową, klienta, procesów wewnętrznych, infrastruktury i rozwoju.

Źródło: Opracowanie własne.

W ostatnich latach nastąpił wzrost zainteresowania chatbotami, a co za tym idzie wzrost liczby publikacji naukowych z nimi związanych. Pomimo większej liczby artykułów, nie wyróżnił się dotychczas jednorodny system metody oceny działania chatbotów. Chatboty najczęściej były oceniane na podstawie testu Turinga. Jednak ten test w ostatnich latach nie cieszy się już taką popularnością, ponieważ w interakcji z chatbotem ludziom nie zależy, aby zdał on ten test.⁸⁴

W literaturze istnieje wiele podejść badawczych dotyczących oceny użyteczności chatbotów. Jednym z nich jest podział na „glass box” oraz „black box”. Badania typu „glass box”

⁸⁴ Casas J., Tricott M-O, Khaled O.A, Mugellini E., Cudre-Maroux P. A.B.. Trends & Methods in Chatbot Evaluation. ICMI '20 Companion, Holandia 2020.

skupiają się na ocenie aspektów funkcjonowania chatbotów takich jak gramatyka, poprawności odpowiedzi, rozumienie języka naturalnego. W drugim podejściu badawczym typu „black box” większą uwagę przykładą się do oceny satysfakcji użytkownika, na którą składa się czas oraz forma usługi. Do tego badania używa się dwóch kwestionariuszy. „Kwestionariusze pokonwersacyjne” wypełniane są od razu po zakończonej konwersacji przez testerów i nie badają dogłębnie satysfakcji. „Kwestionariusze po badaniu” zawierają bardziej wnikliwe oceny użyteczności.⁸⁵

Do ewaluacji chatbotów firmy stosują różne standardy oceny. Jednak znalezienie uniwersalnej miary jest istotne w celu pomiaru wdrożeń najnowszych technologii w systemach informatycznych. W celu bardziej wnikliwego przyjrzenia się miarom oceny chatbotów został przeanalizowany przegląd literatury w tym obszarze⁸⁶.

PARADISE

Kryterium, które dzieli podejścia badawcze, jest podział miar na obiektywne i subiektywne. Model, który bierze pod uwagę obydwa kryteria oceny to PARADISE. PARADISE (PARAdigm for DIAlogue System Evaluation) to zbiór ram oceny agentów opartych na języku mówionym.⁸⁷ PARADISE jest alternatywą testu Turinga uwzględniającą przeznaczenie oraz realizację zadań chatbota. Uwzględnia on obiektywne miary oceny takie jak: procent poprawnych odpowiedzi, jakość rozwiązania, średni czas reakcji systemu, czas dialogu i inne. Miary obiektywne są łatwe dostępne i firma może pozyskać jako dane z systemu. Do subiektywnych miar zalicza się miary intuicyjności interfejsu: łatwość użytkowania, spójność wypowiedzi, poprawność odpowiedzi, zdolność do naprawiania błędów, satysfakcja użytkownika. Dostęp do wymienionych miar jest bardziej skomplikowany i najczęściej wymaga udziału użytkownika po rozmowie z agentem.

AICSQ

Chatboty oparte o moduł przetwarzania języka naturalnego obsługują klientów, których pierwsze wrażenie kontaktu z firmą znacząco wpływa na satysfakcję i lojalność klientów, a co za tym idzie na zyski organizacji. Chatboty, w porównaniu z tradycyjną obsługą klienta,

⁸⁵ Jaskowska M., Wójcik M. Metody i techniki badania użyteczności systemów informatyczno-wyszukiwawczych opartych o model konwersacyjny. Kraków 2017, s.192.

⁸⁶ Casas J., Tricott M-O, Khaled O.A, Mugellini E., Cudre-Maroux P. A.B.. Trends & Methods in Chatbot Evaluation. ICMI '20 Companion, Holandia 2020.

⁸⁷ Jaskowska M., Wójcik M. Metody i techniki badania użyteczności systemów informatyczno-wyszukiwawczych opartych o model konwersacyjny. Kraków 2017, s.192.

opierają się na zaawansowanym systemie informatycznym. Chatboty poprzez zaawansowaną technologię i sztuczną inteligencję gwarantują klientom natychmiastowe odpowiedzi i usługi. Chatboty wyręczają ludzi w pewnych aspektach, do których można zaliczyć przechowywanie informacji, moc obliczeniową i nieograniczoną tempem możliwość ciągłego uczenia się.⁸⁸ Z drugiej strony, chatboty różnią się od systemów informatycznych posiadaniem inteligencji zbudowanej na kształt tej ludzkiej. Dlatego ewaluacja jakości ich usług powinna się różnić od tradycyjnej oceny obsługi klienta. Chen i inni w celu wypełnienia tej luki rozwinęli skalę jakości usług chatbota AICSQ (Artificial Intelligence Chatbot Service Quality), która uwzględnia w ocenie sztuczną inteligencję. Rozszerzona miara AICSQ ma być instrumentem pomiarowym dla empirycznych badań usługi chatbotów, takich jak jakość usług chatbota AI czy mierzenia wyników biznesowych.

W wyniku badań i analizy literatury Chen i inni opracowali model AICSQ pokazany w Tabeli 10. Wymiar jakości zawiera kategorie takie jak: zrozumienie kontekstu, współpraca człowieka ze sztuczną inteligencją, ludzkość, ciągły rozwój, personalizacja, adaptacja kulturowa oraz efektywność.

Tabela 10. Wielopoziomowy wymiar jakości chatbota AI.

Wymiar jakości	Cechy
Zrozumienie kontekstu	• Rozumienie zapytania • Rozumienie wyrażonych emocji
Współpraca człowiek-sztuczna inteligencja	• Dostępność • Łatwość przeniesienia • Zapamiętywanie
Ludzkość	• Wyrażanie ludzkich sygnałów • Ludzka osobowość • Ludzka empatia
Ciągły rozwój	• Samokształcenie • Aktualizacje systemu

⁸⁸ Chen Q., Gong Y., Lu Y., Tang J. Classifying and measuring the service quality of AI chatbot in frontline service. Wuhan 2022, s.552-553.

Personalizacja	• Identyfikacja klienta • Spersonalizowane odpowiedzi • Spersonalizowane polecenia
Adaptacja kulturowa	• Brak barier językowych • Obycie z kulturą
Efektywność	• Ciągła dostępność • Elastyczność • Uproszczenie procesu

Źródło: Opracowanie własne na podstawie: Chen Q., Gong Y., Lu Y., Tang J. Classifying and measuring the service quality of AI chatbot in frontline service. Wuhan 2022, s.553.

Według Chen i innych wysoki poziom AICSQ gwarantuje realizację celów biznesowych. Uważają, że wyniki ich badań dostarczają przedsiębiorstwom obszernej wiedzy na temat wyboru chatbotów. Przed wdrożeniem takiego agenta, organizacje mogą stwierdzić, czy chcą „zatrudnić” takiego wirtualnego asystenta. Dodatkowo, skala AICSQ obrazuje poziom zadowolenia klientów, którzy skorzystali z usług chatbota co przysłużyłoby się do kontroli jakości usług i wydajności organizacji.

ISO 9241

Za popularną normę, która zapewnia jakość usług, uznawana jest standaryzacja ISO (Międzynarodowa Organizacja Normalizacyjna). Norma ISO 9241, obejmuje ergonomię interakcji między człowiekiem, a systemem. Ergonomia jest dyscypliną naukową, która bada przystosowanie maszyn, środowiska do możliwości człowieka. ISO 9241 zawiera wytyczne, jakie powinny spełniać witryny internetowe skupiające się na stworzeniu przyjaznych interfejsów dla użytkownika.⁸⁹ Jaskowska i Wójcik użyteczność systemów konwersacyjnych definiują zgodnie z normą ISO 9241, uznają ją za miarę wydajności, efektywności i satysfakcji użytkownika. Zgodnie z literaturą przedmiotu za wyróżniki systemu konwersacyjnego przyjęto: język, dialog, status i kontrolę, nawigację, zasób wiedzy, pomoc,

⁸⁹ Cagiltay K., Alacam O., Ocak N, Erdal F. Developing ISO 9241-151 Product Certification Process: Challenges. Ankara 2013. s.1-2.

awatar.⁹⁰ Radziwill i Benton⁹¹ na podstawie artykułów napisanych w latach 1995-2017 zauważyli, że metody oceniające chatboty są zgodne z koncepcją użyteczności ISO 9241. Tak jak Jaskowska i Wójcik, wyróżnili oni również wydajność, efektywność i satysfakcję użytkownika jako główne miary jakości chatbota. Efektywność podzielili oni na funkcjonalność, ewaluowaną poprzez interpretację poleceń i łatwość użycia oraz na ludzkość ocenianą głównie za pomocą testu Turinga. Z kolei zadowolenie podzielono na afekt, etykę i zachowanie oraz dostępność. Afekt mierzy się za pomocą pierwszego wrażenia, udzielania wskazówek i informacji emocjonalnych. Natomiast etyka i zachowanie odnoszą się do godności, etyki oraz kultury, a dostępność dotyczy reakcji na sygnały społeczne i wykrywanie intencji użytkownika.

BLEU i ROUGE

Język używany przez chatbota powinien zawierać rozpoznawalne słowa, frazy i pojęcia, tak aby użytkownik mógł go zrozumieć.⁹² Naturalna rozmowa z chatbotem może być prowadzona poprzez podtrzymywanie tematu, na przykład poprzez zadawanie dotyczących tego tematu pytań.⁹³ Baza wiedzy chatbotów powinna zawierać wiedzę ogólną, jako że ma imitować rozmowę z człowiekiem.⁹⁴ Komunikaty o błędach powinny być dla użytkownika łatwe do rozpoznania i zrozumienia co dokładnie nie działa i co zrobić, aby to naprawić.⁹⁵ Ważną funkcją jest też możliwość dostępu do menu, bądź informacji oraz do złożenia opinii po skorzystaniu z usługi. Chatbot powinien posiadać awatar, na który powinna się złożyć jego osobowość, płeć, imię, wiek oraz wyrażać emocje.⁹⁶

Maroengsit i inni⁹⁷ w 2019 roku w swoim artykule przeanalizowali 30 artykułów, spośród których 17 zawierało metody oceny chatbotów. Ocena treści zawiera automatyczną ocenę za pośrednictwem BLEU i ROUGE. Druga metoda skupia się na satysfakcji użytkownika i w

⁹⁰ Jaskowska M., Wójcik M. Metody i techniki badania użyteczności systemów informatyczno-wyszukiwawczych opartych o model konwersacyjny. Kraków 2017, s.192.

⁹¹ Radziwill N., Benton M. Evaluating Quality of Chatbots and Intelligent Conversational Agents. 2017.

⁹² Scott K. Usability heuristics for bots. 2016. <https://thekevinscott.com/usability-heuristics-for-bots/> [dostęp 08.2022]

⁹³ Morrissey K., Kirakowski J., Realness in chatbots: establishing quantifiable criteria. 2013. s. 87-96.

⁹⁴ Jaskowska M., Wójcik M. Metody i techniki badania użyteczności systemów informatyczno-wyszukiwawczych opartych o model konwersacyjny. Kraków 2017, s.197.

⁹⁵ Scott K. Usability heuristics for bots., op.cit.

⁹⁶ Kuligowska K., Commercial chatbot: performance evaluation, usability metrics, and quality standards of embodied conversational agents, Professionals Center for Business Research, Volume 2 – Feb 2015 (02), Alkhaer Publications, Manchester 2015.

⁹⁷ Maroengsit W., Piyakulpinyo T. Phonyiam K., Pongnumkul S., Chaovalit P., Theeramunkong, T. A Survey on Evaluation Methods for Chatbots. 7th Inter- national Conference on Information and Education Technology. 2019.

bezpośredni sposób pyta użytkowników o zadowolenie z użycia chatbota. Ocena funkcjonowania obejmuje ocenę zadaniową dla chatbotów zadaniowych. W zależności od celów chatbota można zastosować cztery perspektywy ewaluacji. Perspektywa użytkownika mierzy satysfakcję użytkownika i użyteczność. Perspektywa wyszukiwania informacji mierzy dokładność i szybkość odpowiedzi na wiadomość użytkownika. Perspektywa językowa bada w jakim stopniu konwersacja jest jakościowa, ilościowa, a perspektywa AI jest badana przez test Turinga.

HCI

Nadrzędnym celem systemu konwersacyjnego jest usprawnienie relacji użytkownika z przedsiębiorstwem. Używanie systemu konwersacyjnego powinno być intuicyjne i łatwe. Na takim założeniu powinien się opierać system, który bazuje na modelach konwersacyjnych w języku naturalnym.⁹⁸ Rozwój sztucznej inteligencji umożliwia stworzenie właśnie takich systemów, a popularność takich rozwiązań można potwierdzić danymi dotyczącymi licznych wdrożeń w przedsiębiorstwach. Jasińska i Wójcik uznały za konieczne wypracowanie metod i technik diagnozowania dostosowanych do specyfiki takich systemów poprzez badanie ich funkcjonalności, użyteczności i przyjazności dla użytkowników.⁹⁹ Jain i inni¹⁰⁰ skupili się na interakcji człowieka z komputerem (ang. Human Computer Interaction, HCI). Po trzydniowych rozmowach użytkowników z chatbotem zostały zbadane dane ilościowe takie jak czas interakcji, liczba wiadomości. Na koniec okresu użytkownicy ocenili funkcjonowanie chatbota na płaszczyznach takich jak: funkcjonalność, konwersacja, inteligencja, osobowość chatbota i interfejs chatbota. Venkatesh i inni¹⁰¹ na podstawie danych z Alexa Prize, bazy z milionem konwersacji z chatbotami, odrzucają Test Turinga jako sposób oceny chatbotów. Zaproponowany został model wskaźników badających chatbota: doświadczenie konwersacyjne użytkownika, zaangażowanie, spójność, zasięg domeny, głębia konwersacji i różnorodność tematyczna, zakres konwersacji.

Wymienione miary, zebrane poniżej w Tabeli 11, były najczęściej poruszane i opisywane w artykułach naukowych dotyczących miar funkcjonowania chatbotów. Na ich podstawie

⁹⁸ Próchnicka M. Człowiek i komputer. Dialogowy model wyszukiwania informacji. Wydawnictwo UJ. Kraków 2004.

⁹⁹ Jaskowska M., Wójcik M. Metody i techniki badania użyteczności systemów informatyczno-wyszukiwawczych opartych o model konwersacyjny. Kraków 2017, s.197.

¹⁰⁰ Jain M., Kumar P., Kota R., Patel S. Evaluating and Informing the Design of Chatbots. Hong Kong 2018.

¹⁰¹ Venkatesh A., Khatri C., Ram A., Guo F., Gabriel R., Nagar A., Prasad R., Cheng M., Hedayatnia B., Metallinou A., Goel R., Yang S., Raju A. Evaluating and Comparing Conversational Agents. USA 2018.

można zauważyć, że przedstawione miary funkcjonowanie systemów informatycznych są zbieżne do miar ISO 9241, co jest zgodnie z wnioskami na podstawie przeprowadzonej już przeglądu literatury przedmiotu. Warto zauważyć, że popularny test Turinga dalej jest stosowany, ale nie jest on już samodzielną miarą oceny chatbotów. Jest on najczęściej używany jedynie jako narzędzie do pomiaru jednej z perspektyw używanych do oceny efektywności systemu informatycznego opartego o model konwersacyjny. Chatboty ewaluowane są również pod względem jakości obsługi klienta. Z kolei miara AICSQ składa się z dodatkowych kryteriów oceny.

Tabela 11. Podsumowanie wybranych podejść dotyczących oceny miar funkcjonowania chatbotów.

Podejście	Najważniejsze informacje
ISO 9241	Norma oceny ergonomii interakcji człowiek – komputer. Kryteria oceny: 1. wydajność, 2. efektywność, 3. satysfakcja. Wyróżniki systemu przyjęte przez Jaskowską i Wójcik to: język, dialog, status i kontrolę, nawigację, zasób wiedzy, pomoc, i awatar.
AICSQ	Jakość usług chatbota AICSQ (Artificial Intelligence Chatbot Service Quality) mierzona jest przy użyciu miar: 1. zrozumienie kontekstu, 2. współpraca człowiek – sztuczna inteligencja, 3. ludzkość, 4. ciągły rozwój, 5. personalizacja, adaptacja kulturowa, efektywność.
PARADISE	Zbiór obiektywnych i subiektywnych ram ocen. Do obiektywnych zalicza się m.in.: 1. procent poprawnych odpowiedzi, 2. średni czas reakcji systemu, 3. jakość rozwiązania. Do subiektywnych należą miary takie jak:

	1. łatwość użytkowania, 2. spójność wypowiedzi, 3. satysfakcja użytkownika.
--	---

Źródło: opracowanie własne

System oceny Russel-Rose

Do ekonomicznych miar oceny chatbotów można zaliczyć system oceny, który początkowo został stworzony przez Russel-Rose'a. Russel-Rose użył czterech perspektyw służących do oceny funkcjonowania chatbotów. Do miar oceny zaliczył doświadczenie użytkownika, wyszukiwanie informacji, perspektywę językową oraz perspektywę technologiczną. Podejście Russel-Rose'a zostało jednak rozszerzone przez Dijanę Peras, która przedstawiła dodatkową perspektywę chatbotów. Za piątą perspektywę uznała perspektywę biznesową, która objęła wartości biznesowe i ekonomiczne, dotychczas nieuwzględnione w tym podejściu.¹⁰² Na podstawie perspektyw stworzono ramy oceny, które składają się z kategorii, atrybutów, metryk oraz podejść.

Pierwszą perspektywą jest perspektywa doświadczeń użytkownika. Jest ona złożona z czterech kategorii takich jak użyteczność, wydajność, afekt oraz satysfakcja. Użyteczność wiąże się z łatwością obsługi i czasem potrzebnym na wykonanie zadania. Na jakość wykonania składa się kompletność, terminowość i inne jakościowe charakterystyki. Na afekt ma wpływ ocena doświadczeń i emocji korzystających z chatbota użytkowników. Afekt mierzony jest za pomocą wypełnianych ankiet bądź kwestionariuszy. Satysfakcja mierzy zadowolenia osoby korzystającej z usług chatbota. Porównuje się oczekiwane odczucie z faktycznym odczuciem na temat skorzystania z chatbota.

Kolejną perspektywą jest perspektywa wyszukiwania informacji, która weryfikuje zaspokajanie potrzeby informacji przez firmy. Składa się ona z trzech kategorii: dokładności, dostępności, efektywności. Te kategorie mają charakter ilościowy i badają w jakim stopniu realizowane jest dane zadanie i dostarczają wartościowych informacji. Natomiast perspektywa językowa weryfikuje poziom poprawności językowej chatbotów i ich odpowiedzi. Wydajność natomiast mierzona jest za pomocą poziomu słowno-gramatycznego. Perspektywa językowa składa się z kategorii takich jak: jakość, ilość, relacja, sposób,

¹⁰² Peras D., Chatbot evaluation metrics: review paper, Economic and Social Development, Zagrzeb 2018.

poprawność gramatyczna. Pierwsze cztery sprawdza się za pomocą skali Likerta. Wymienione miary tej perspektywy mają charakter jakościowy i wymagany jest zespół ekspertów do przeprowadzenia takiej oceny.

Do perspektywy technologii zalicza się jedną kategorię: „ludzkość”. Chatbot powinien mieć zdolność do zachowywania i wyrażania ludzkich zachowań. Konieczne jest więc przetwarzanie, rozumienie i generowanie języka naturalnego (NLP, NLU, NLG). Dodatkowo, „ludzkość” wirtualnego asystenta najczęściej mierzy się miarą ilościową poprzez test Turinga. Miarą jakościową wykorzystywaną jest między innymi skala ocen, chociaż jest to subiektywna ocena. Ostatnio wprowadzoną perspektywą jest perspektywa biznesowa, która jest kluczowa dla przedsiębiorstw, jako wypadkowa skuteczności i kosztów chatbota. Pomiar skuteczności może być dokonywany za pomocą wskaźników takich jak: liczba użytkowników, czas rozmowy chatbota, liczba rozmów chatbota. Natomiast dla oszacowania kosztów bierze się pod uwagę: liczbę agentów prowadzących konwersację, czas trwania rozmowy, liczbę nieudanych konwersacji, liczbę nietrafnych odpowiedzi udzielonych przez chatbota.

Literatura przedmiotu rzadko skupia się na ekonomicznych miarach wdrożenia chatbotów w przedsiębiorstwach. Większą popularnością cieszą się miary jakości systemu, które nie skupiają się na korzyściach i kosztach. Dlatego oprócz wskaźników charakterystycznych dla klasycznych systemów informatycznych, chatboty powinny być oceniane poprzez miary, które dotyczą jakości obsługi klienta oraz weryfikują adekwatny poziom działania sztucznej inteligencji.

Spośród niewielkiej liczby artykułów podejmujących przegląd ekonomicznych aspektów funkcjonowania chatbotów, na uwagę zasługuje raport firmowy IBM, w którym posłużono się miarami i aspektami funkcjonowania klasycznych systemów informatycznych. Raport przeprowadzony został przez firmę Forrester, która to stworzyła miarę TEI służącą do ekonomicznej oceny systemów informatycznych.¹⁰³ Firma Forrester, używając między innymi metody całkowitego wpływu ekonomicznego, wyliczyła koszty i korzyści biznesowe zaoszczędzone dzięki wdrożeniu w firmie IBM (International Business Machines

¹⁰³ Forrester Total Economic Impact. The Total Economic Impact of IBM Watson Assistant 2020, https://www.ibm.com/watson/assets/duo/pdf/watson_assistant/The_Total_Economic_Impact_of_IBM_Watson_Assistant-March_2020_v3.pdf [dostęp 08.2022]

Corporation) chatbota o nazwie Watson Assistant. Celem raportu jest dostarczenie ram oceny możliwego wpływu finansowego chatbota Watson Assistant do organizacji. Watson Assistant jest chatbotem firmy IBM, który przy użyciu przetwarzania języka naturalnego umożliwia użytkownikowi interakcje z systemem biznesowym. Chatbot w firmie IBM zastąpił usługi chatu, e-maile i rozmowy telefoniczne, gdyż te metody uznano za zbyt wolne i kosztowne.

Firmy decydują się na inwestycję w chatbota Watson Assistant, aby wydane koszty przełożyły się na bardziej efektywne działania oraz w celu poprawienia jakości obsługi klienta. Watson Assistant jest pierwszym poziomem zetknięcia się klienta z przedsiębiorstwem za pośrednictwem komunikatorów. Watson umie odpowiadać na proste pytania użytkowników, pomaga klientom znaleźć potrzebne dla nich informacje oraz przekierowuje użytkownika do ludzkiego asystenta w przypadku bardziej skomplikowanego zadania.

Sukces wirtualnego asystenta Watson Assistant mierzony jest na podstawie korzyści wynikających z oszczędności związanych z konwersacją, konsolidacją helpdesku, agentów IT i zasobów ludzkich, wzrostem wydajności dzięki obecności agenta, prawidłowym przekierowaniu rozmów do pracownika. W zamieszczonej dalej Tabeli 12. przedstawione zostały dokładne wyniki uzyskane po wdrożeniu agenta do firmy IBM. Suma wszystkich korzyści wynosi prawie 24 miliony dolarów z uwzględnieniem wartości bieżącej skorygowanej o ryzyko.

Tabela 12. Korzyści z wdrożenia chatbota Watson Assistant.

Korzyść	ROK 1	ROK 2	ROK 3	SUMA	Wartość bieżąca skorygowana o ryzyko
Oszczędności związane z konwersacją z klientem	\$1,584,000	\$4,989,600	\$9,900,000	\$16,473,600	\$13,001,663
Konsolidacja helpdesku, agentów IT i zasobów ludzkich	\$0	\$2,040,000	\$2,040,000	\$4,080,000	\$3,218,633
Wzrost wydajności dzięki obecności agenta	\$0	\$0	\$1,350,000	\$1,350,000	\$1,014,275
Prawidłowe	\$2,276,640	\$2,656,080	\$3,225,240	\$8,157,960	\$6,687,951

przekierowanie rozmowy do pracownika					
Suma	\$3,860,640	\$9,685,680	\$16,515,240	\$30,061,560	\$23,922,512

Źródło: Opracowanie własne na podstawie: Forrester Total Economic Impact. The Total Economic Impact of IBM Watson Assistant 2020,
https://www.ibm.com/watson/assets/duo/pdf/watson_assistant/The_Total_Economic_Impact_of_IBM_Watson_Assistant-March_2020_v3.pdf [dostęp 08.2022].

Do kosztów chatbota Watson Assistant wliczono jego licencję, wewnętrzne koszty pracy, płace analityków konwersacyjnych, opłaty za usługi. Zamieszczona dalej Tabela 13. obrazuje podział kosztów wraz z poniesionymi opłatami. Suma kosztów w okresie trzech lat od implementacji wyniosła blisko 5,5 miliona dolarów.

Tabela 13. Koszty wdrożenia chatbota Watson Assistant.

Koszty	Początkowe	Rok 1	Rok 2	Rok 3	Suma	Wartość bieżąca skorygowana o ryzyko
Licencja IBM	\$0	\$20,700	\$80,213	\$138,000	\$238,913	\$188,791
Wewnętrzne koszty pracy	\$14,200	\$93,150	\$62,100	\$62,100	\$341,550	\$306,861
Analitycy konwersacji	\$742,500	\$742,500	\$1,485,000	\$2,227,500	\$5,197,500	\$4,318,326
Opłaty za usługi	\$287,500	\$120,750	\$208,150	\$121,900	\$738,300	\$660,883
Suma kosztów	\$1,154,200	\$997,100	\$1,835,463	\$2,549,500	\$6,516,263	\$5,474,861

Źródło: Opracowanie własne na podstawie: Forrester Total Economic Impact. The Total Economic Impact of IBM Watson Assistant 2020,
https://www.ibm.com/watson/assets/duo/pdf/watson_assistant/The_Total_Economic_Impact_of_IBM_Watson_Assistant-March_2020_v3.pdf [dostęp 08.2022].

Dla wdrożenia chatbota Watson Assistant wartość współczynnika NPV wyniosła 18,4 miliona dolarów, wartość współczynnika ROI wyniosła 337%, a czas zwrotu z inwestycji wyniósł mniej niż pół roku. Firma Forrester do analizy wpływu chatbota Watson Assistant użyła podstawowych elementów TEI: korzyści, kosztów, elastyczności i ryzyka.

Rozdział 6. Porównanie miar oceny wdrożenia systemów informatycznych niekonwersacyjnych oraz opartych o model konwersacyjny

Ocena chatbotów opartych o przetwarzanie języka naturalnego łączy kilka miar ich funkcjonowania w przedsiębiorstwie. Jednocześnie chatboty powinny podlegać miarom, które dotyczą jakości obsługi klienta, miarom jakim podlegają systemy informatyczne oraz tym, które weryfikują poziom sztucznej inteligencji. Chatboty powinny podlegać szerszej ocenie niż tradycyjna obsługa klienta, ponieważ nie uwzględnia ona technicznych cech usługi chatbotów.¹⁰⁴ Poziom sztucznej inteligencji mierzony jest za pomocą testu Turinga. Oprócz tych miar, chatboty podlegają ocenie tak, jak klasyczne systemy informatyczne i mierzy się ich korzyści i koszty wdrożenia i funkcjonowania.

Jakość klasycznych systemów informatycznych mierzona jest za pomocą normy jakości ISO 25010:2011. Jest ona uważana za międzynarodową normę oprogramowania. Jednak zgodnie z przytoczoną literaturą przedmiotu chatboty podlegają innej normie jakości. Tą normą jest ISO 9241, która przede wszystkim zawiera wytyczne do stworzenia przyjaznych interfejsów dla użytkownika stron internetowych. W zamieszczonej dalej Tabeli 14. zostały zestawione wymienione obydwie normy ISO. Na pierwszy rzut oka widać dużą rozbieżność w ocenie systemów. Klasyczne systemy informatyczne posiadają szerszą perspektywę oceny. Chatboty ewaluowane są przede wszystkim pod względem dostosowania do użytkownika i jego zadowolenia.

Tabela 14. Porównanie występujących w literaturze przedmiotu ekonomicznych miar implementacji i funkcjonowania systemów informatycznych niekonwersacyjnych i systemów informatycznych opartych o model konwersacyjny.

Norma ISO/IEC 25010:2011	Norma ISO 9241
Norma oceny jakości oprogramowania zawiera: model jakości, miary zewnętrzne, miary wewnętrzne, miary jakości użytkowej. Kryteria oceny: 1. funkcjonalność,	Norma oceny ergonomii interakcji człowiek – komputer. Kryteria oceny: 1. wydajność, 2. efektywność, 3. satysfakcja.

¹⁰⁴ Chen Q., Gong Y., Lu Y., Tang J. Classifying and measuring the service quality of AI chatbot in frontline service. Wuhan 2022, s.552-553.

2. niezawodność, 3. użyteczność, 4. wydajność, 5. możliwość konserwacji, 6. przenośność, 7. bezpieczeństwo, 8. zgodność	Wyróżniki systemu przyjęte przez Jaskowska i Wójcik: język, dialog, status i kontrolę, nawigację, zasób wiedzy, pomoc, awatar
---	---

Źródło: Opracowanie własne.

Porównanie ekonomicznych aspektów implementacji klasycznych systemów informatycznych i systemów informatycznych opartych o model konwersacyjny nie jest sprawą prostą. W przeciwieństwie do szeroko opisanych miar oceny funkcjonowania klasycznych systemów informatycznych, systemy oparte o model konwersacyjny nie doczekały się wielu artykułów, które podejmują się tej tematyki. Jednak dzięki dostępnemu raportowi przeprowadzonemu przez firmę Forrester można założyć, że pomimo iż prezentowane tam miary nie były zbadane w artykułach naukowych, to można stosować je również w przypadku chatbotów.

W literaturze dla klasycznych systemów informatycznych można znaleźć ekonomiczne miary oceny takie jak: NPV (Net Present Value) – bieżąca wartość netto, ROI (Return of Investment) - zwrot z inwestycji, TCO (Total Cost of Ownership) - metoda pełnych kosztów posiadania, TEI (Total Economic Impact) - metoda całkowitego wpływu ekonomicznego, Ekonomia Informacji, BS (Balanced Scorecard) - metoda zrównoważonej karty wyników.

W przypadku systemów opartych o model konwersacyjny można znaleźć miarę składającą się z pięciu perspektyw, wewnątrz której jest też perspektywa biznesowa. Tę metodę można porównać z metodą zrównoważonej karty wyników, gdyż obie rozpatrują perspektywy, które grupują mierzone cele systemu. Ekonomia Informacji nie znalazła swojego odzwierciedlenia w literaturze zajmującej się chatbotami. Miary ROI, TCO, TEI oraz NPV nie zostały opisane w literaturze chatbotów, jednak zgodnie z raportem firmy Forrester, powyższe miary były zastosowane do obliczenia wyniku ekonomicznego wdrożenia Watson Assistant. W zamieszczonej dalej Tabeli 15. przedstawiono omówione ekonomiczne miary implementacji i funkcjonowania porównywanych systemów.

Tabela 15. Porównanie ekonomicznych miar implementacji i funkcjonowania systemów i systemów informatycznych opartych o model konwersacyjny.

Ekonomiczne miary implementacji	Systemy informatyczne	
	klasyczne	oparte o model konwersacyjny
NPV	+	+
ROI	+	+
TCO	+	+
TEI	+	+
Ekonomika Informacji	+	?
BS	+	?
5 perspektyw doświadczeń użytkownika	?	+

Źródło: Opracowanie własne.

W literaturze przedmiotu miary oceny implementacji chatbotów skupiają się przede wszystkim na ocenie jakości. Składa się na nią najczęściej satysfakcja użytkownika, sprawność chatbota, perspektywa językowej i test Turinga. Można założyć, iż motywacją do wdrożenia chatbota dla firmy może być podążanie za trendami informatycznymi, traktowanie chatbotów jako element unowocześnienia organizacji i „bycia na czasie”. Firmy w takiej ocenie chcą sprawdzić czy ich chatbot przede wszystkich spełnia wymogi klientów.

Brak artykułów naukowych podejmujących temat ekonomicznych aspektów systemów informatycznych opartych o model konwersacyjny wynika z faktu, że wciąż mało jest artykułów i opracowań w temacie samych chatbotów. Natomiast przyczyną możliwej przewagi w stosowaniu miar jakościowych (względem ekonomicznych) dla chatbotów jest przykładanie do systemów informatycznych opartych o model konwersacyjny podobnych miar kosztów i korzyści, jak do klasycznych systemów informatycznych.

Zakładając powyższy scenariusz, firmy do zbadania skuteczności chatbota pod względem ekonomicznym użyją miar typu ROI, TCO, TEI. Takie możliwe podejście potwierdza przytoczony raport firmy Forrester¹⁰⁵ sporządzony dla IBM. Natomiast ocena jakości takiego systemu bardziej odbiega od tradycyjnej oceny systemów informatycznych. Ocena chatbotów różni się koniecznością zwrócenia większej uwagi na aspekt sztucznej inteligencji oraz „ludzkość” systemu. Jest to system nastawiony na komunikację, który w największym stopniu

¹⁰⁵ Forrester Total Economic Impact. The Total Economic Impact of IBM Watson Assistant 2020, https://www.ibm.com/watson/assets/duo/pdf/watson_assistant/The_Total_Economic_Impact_of_IBM_Watson_Assistant-March_2020_v3.pdf [dostęp 08.2022]

powinien być również pod tym względem ewaluowany. Przedsiębiorstwa zazwyczaj inaczej mierzą funkcjonowanie i wdrożenie chatbotów niż jest to przyjęte dla klasycznych systemów informatycznych, zgodnie z teorią nauki informatyki gospodarczej.

W omawianych miarach główną miarą jakości klasycznych systemów informatycznych okazała się norma oceny jakości oprogramowania ISO/IEC 25010:2011. System oceniany jest na podstawie normy złożonej z ośmiu kategorii, w skład których wchodzi: funkcjonalność, niezawodność, użyteczność, wydajność, możliwość konserwacji, przenośność, bezpieczeństwo i zgodność. Natomiast chatboty podlegają innym normom, nawet pod względem normy jakości. Większość miar jakości jest zbliżona do norm oceny ISO 9241. Jest to norma, która skupia się na ergonomii komunikacji człowieka z komputerem i wpływa na nią: wydajność, efektywność, satysfakcja podczas interakcji człowieka z komputerem. Jest to więc norma skupiająca się na innych aspektach oceny jakości niż w przypadku klasycznych systemów informatycznych. Również odmienne spojrzenie na chatboty jest zgodne z podejściem AICSQ. Jakość usługi chatbota zależy od jego zrozumienia kontekstu wypowiedzi, owocnej współpracy z użytkownikiem, naśladowania ludzkich zachowań, rozwoju i zdolności adaptacji.

Klasyczne systemy informatyczne doczekały się wielu ekonomicznych miar oceniających aspekty wdrożenia i funkcjonowania systemów informatycznych. Do ekonomicznych miar zalicza się miary, które bazują na korzyściach i kosztach, takich jak: zwrot z inwestycji, metoda pełnych kosztów posiadania, metoda całkowitego wpływu ekonomicznego czy bieżąca wartość netto. Dla wewnętrznej zbilansowanej karty wyników i ekonomiki informacji perspektywa biznesowa jest jedną z ocenianych perspektyw. Metody ekonomiczne grupują między innymi koszty, które mają pomóc w identyfikacji ukrytych wydatków. Ich zadaniem jest też oszacowanie opłacalności projektu, metoda zwrotu z inwestycji bazuje na zmianie w przychodach, kosztach, korzyściach niemierzalnych, kosztach nabycia systemu informatycznego i innych. Systemy informatyczne mogą być ewaluowane przy użyciu ekonomicznych miar wykorzystujących: bezpośrednie, pośrednie, finansowe, niefinansowe koszty i korzyści

Literatura przedmiotu nie dostarcza jednak wielu ekonomicznych miar oceny funkcjonowania chatbotów. Miara Russel-Rose'a zaktualizowana przez Dijanę Peras składa się z pięciu perspektyw, w tym z perspektywy biznesowej. Wartość biznesowa mierzona jest poprzez

różnice pomiędzy korzyściami, a kosztami chatbota. Skuteczność uwzględnia: liczbę użytkowników, czas rozmowy czy liczbę przeprowadzonych rozmów. Natomiast do kosztów zalicza się ilość użytych zasobów do rozmowy, czas jej trwania, liczbę pomyłek chatbota i udzielonych niepoprawnych odpowiedzi. Firmy często wykupują licencję na chatbota, która oprócz opłaty początkowej, pobiera każdorazowo opłatę za rozmowę przeprowadzoną przez chatbota. Kolejnym źródłem, które może pomóc w sklasyfikowaniu ekonomicznych miar implementacji jest wspomniany raport firmy Forrester dla firmy IBM. W raporcie firmy Forrester¹⁰⁶ oszacowano, że koszt każdorazowej rozmowy wirtualnego asystenta wyceniany jest na 0,5\$, przy 6\$ dla pracownika. Wirtualny asystent Watson został poddany ekonomicznym miarom oceny, takim jak klasyczne systemy informatyczne. Wskaźniki Watsona zostały spersonalizowane dla potrzeb chatbota i użyto odmiennych kategorii kosztów i korzyści.

¹⁰⁶ Forrester Total Economic Impact. The Total Economic Impact of IBM Watson Assistant 2020, https://www.ibm.com/watson/assets/duo/pdf/watson_assistant/The_Total_Economic_Impact_of_IBM_Watson_Assistant-March_2020_v3.pdf [dostęp 08.2022]