CASIE i pokrewne nurty badań nad kontekstowym rozpoznawaniem emocji

W przeszukanej literaturze naukowej akronim **CASIE** nie funkcjonuje przede wszystkim jako „Contextual Affect Sensing for Intelligent Environments", lecz jako **Computing Affect and Social IntelligencE** - nazwa zaproponowanej w 2021 roku architektury dla robotów społecznych w ochronie zdrowia. Równolegle istnieje wcześniejszy, blisko spokrewniony nurt badań nad **context-sensitive / contextual affect sensing**, rozwijany głównie przez **Li Zhang** i współpracowników w obszarze agentów konwersacyjnych i wirtualnej improwizacji dramatycznej, oraz pokrewny nurt **context-aware emotional intelligence** rozwijany przez **Michala Ptaszyńskiego** i współautorów. Z tego powodu najbardziej rzetelne ujęcie „frameworków CASIE" to nie jedna ustalona rodzina modeli, lecz **trzy powiązane linie architektoniczne**: kontekstowe wykrywanie afektu w dialogu, kontekstową ocenę adekwatności emocji oraz pełniejszą architekturę robotyczną CASIE dla opieki zdrowotnej.

Najważniejszy wniosek architektoniczny jest taki, że **CASIE nie jest klasyfikatorem „emocja z twarzy = reakcja"**. W wersji robotycznej rozszerza klasyczną pętlę *sense–think–act* o: multimodalną percepcję mowy i obrazu, **fuzję multimodalną**, **bazę wiedzy / knowledge graph**, **High-Level Control Centre**, **emotion-based decision-making**, **probabilistyczny planner** oraz **normatywny komponent etyczno-społeczny**. To odróżnia CASIE od czysto korelacyjnych systemów rozpoznawania emocji [2].

Jednocześnie CASIE - w opublikowanej postaci - **nie dostarcza jeszcze jawnego formalizmu przyczynowego** w sensie współczesnych modeli SCM/counterfactual reasoning, ani nie rozwija kompletnego mechanizmu **kalibracji niepewności, abstencji czy proszenia użytkownika o doprecyzowanie**. W tym sensie bliżej mu do **hybrydowego rozumowania logiczno-probabilistycznego** niż do pełnego modelu przyczynowego. Współczesne prace z obszaru context-aware emotion recognition, takie jak **CLEF** z CVPR 2024, zaczynają już explicite modelować graf przyczynowy i odróżniać użyteczny kontekst od kontekstu będącego źródłem pozornych korelacji. Osobny nurt wnosi formalne modelowanie niepewności, np. **HU-Dialogue** oraz **conformal prediction** dla ERC [3].

Praktycznie rzecz ujmując, CASIE jest dziś najlepiej czytać jako **program badawczy i architekturę referencyjną**, a nie jako skończony, zwalidowany end-to-end framework produkcyjny. Jego mocną stroną jest bogate ujęcie kontekstu, planowania i etyki; słabszą - brak pełnej operacjonalizacji przyczynowości, niepewności i polityk „nie wiem / zapytaj człowieka". Dla współczesnych zastosowań należałoby więc budować **CASIE-2**: z jawnym stanem przekonań, kalibracją, mechanizmem aktywnego dopytywania i restrykcyjną polityką bezpieczeństwa działań [4].

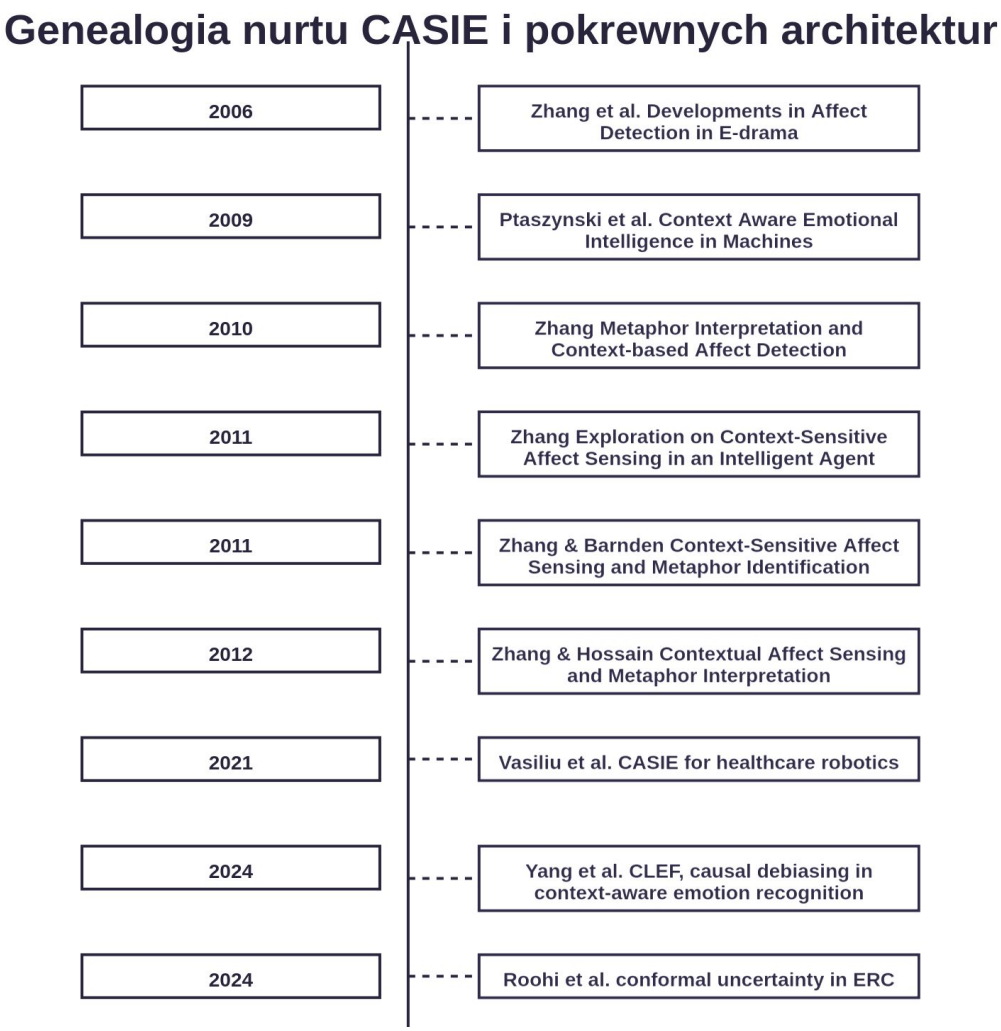
Co w literaturze naprawdę oznacza CASIE

W recenzowanych źródłach, które udało się zidentyfikować, **CASIE** jako rozwinięty akronim pojawia się przede wszystkim w artykule **Vasiliu i in.** w *Paladyn, Journal of Behavioral Robotics* jako **Computing Affect and Social IntelligencE** dla robotów społecznych w opiece zdrowotnej. Autorzy explicite opisują CASIE jako „emotion-aware architecture" łączącą affective computing, NLP, computer vision i complex decision-making in HRI. Repozytorium DCU oznacza ten artykuł recenzowany [5].

Starsza literatura, z którą CASIE jest ideowo spójne, ale która **nie używa jeszcze tego akronimu**, mówi o **context-sensitive affect sensing**, **contextual affect sensing** oraz **context-aware emotional intelligence**. U **Li Zhang** rdzeniem problemu jest interpretowanie afektu z wypowiedzi, które mają słabe lub zerowe jawne wskaźniki emocjonalne, lecz nabierają znaczenia dzięki **historii dialogu, relacjom między postaciami, tematowi rozmowy i modelowaniu nastroju postaci**. U **Ptaszyńskiego i współautorów** celem jest krok dalej: nie tylko rozpoznać emocję, ale też ocenić, czy jest ona **adekwatna do kontekstu**, aby agent potrafił wybrać właściwą strategię odpowiedzi.

Tak więc, jeśli przez „frameworki CASIE" rozumieć **szerszą rodzinę kontekstowo-rozumujących systemów afektywnych**, to najbardziej trafny podział wygląda następująco:

1. po pierwsze, **kontekstowe wykrywanie afektu w dialogu**;
2. po drugie, **kontekstowa ocena adekwatności afektu**;
3. po trzecie, **pełna architektura robotyczna CASIE**, w której percepcja afektu jest tylko jedną z warstw całego systemu [7].



Oś czasu pokazuje, że CASIE z 2021 roku nie powstało w próżni: wyrasta z co najmniej piętnastoletniej tradycji badań nad tym, jak **kontekst przekształca znaczenie sygnałów emocjonalnych** i jak agent może działać adekwatnie społecznie, a nie tylko reagować na korelacje w danych [8].

Kluczowe publikacje i miejsca publikacji

Rok	Publikacja	Autorzy	Znaczenie architektoniczne
2006	Developments in Affect Detection in E-drama	Li Zhang, John A. Barnden, Robert J. Hendley, Alan M. Wallington; **EACL Demonstrations**	Wczesny system wykrywania afektu w e-dramie; punkt wyjścia dla późniejszego modelowania kontekstu.
2009	Towards Context Aware Emotional Intelligence in Machines	Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka, Kenji Araki; **IJCAI**	Wprowadza pojęcie **adekwatności emocji względem kontekstu** i agentowej strategii reakcji.
2010	Metaphor Interpretation and Context-based Affect Detection	Li Zhang; **COLING 2010 Poster Volume**	Łączy interpretację metafor z kontekstowym wykrywaniem afektu i modelowaniem Markowowskim.
2011	Exploration on Context-Sensitive Affect Sensing in an Intelligent Agent	Li Zhang; **IVA 2011**	Dodaje wnioskowanie o nastroju improwizacyjnym postaci i emocjach osadzonych w kontekście komunikacyjnym.
2011	Context-Sensitive Affect Sensing and Metaphor Identification in Virtual Drama	Li Zhang, John Barnden; **ACII 2011**	Formalizuje połączenie kontekstu i metafory w architekturze wirtualnego dramatu.
2012	Contextual affect sensing and metaphor interpretation	Li Zhang, Alamgir Hossain; **Entertainment Computing**	Artykuł journalowy porządkujący nurt; opisuje profil kontekstu, pewność odpowiedzi i wyniki ewaluacji.
2021	CASIE – Computing affect and social intelligence for healthcare in an ethical and trustworthy manner	Laurentiu Vasiliu i in.; **Paladyn, Journal of Behavioral Robotics**	Najpełniejsza architektura CASIE: multimodalna percepcja, knowledge base, planner, etyka, edge/cloud.
2024	Robust Emotion Recognition in Context Debiasing	Dingkang Yang i in.; **CVPR 2024**	Nowoczesny punkt odniesienia: jawny **graf przyczynowy** i kontrfaktyczne usuwanie biasu kontekstowego.

Genealogia i architektury poprzedzające CASIE

Najwcześniejsza linia, związana z projektem **EMMA / e-drama**, startowała od rozpoznawania afektu bezpośrednio z każdej pojedynczej wypowiedzi użytkownika. Już wtedy jednak autorzy zauważyli, że wiele emocjonalnie ważnych wypowiedzi nie zawiera „silnych wskaźników afektu", a ich znaczenie ujawnia dopiero kontekst. W pracy z COLING 2010 Li Zhang opisuje użycie **analizy kontekstu językowego**, **poznawczego modelowania emocji opartego na łańcuchach Markowa** oraz „dynamicznego algorytmu" do interpretacji afektu z kontekstu w otwartym dialogu tekstowym. Ewaluacja na transkryptach scenariusza szkolnego bullyingu dała około **69% średniej dokładności** dla kontekstowego interpretowania emocji postaci [11].

W dojrzałej wersji tego nurtu, opisanej w *Entertainment Computing*, moduł kontekstu obejmował **improwizacyjny nastrój danej postaci**, **ostatnie profile emocjonalne innych postaci**, **wątek tematyczny rozmowy** oraz detekcję emocji z wyrażeń metaforycznych. Co ważne, agent nie reagował w sposób bezwzględny: autorzy wskazują, że **AI agent może dostosować prawdopodobieństwo odpowiedzi do tego, jak pewny jest swojej interpretacji**, co można uznać za wczesny, jeszcze nieusystematyzowany odpowiednik epistemicznej pokory. W tej samej pracy odnotowano m.

in. **68%** trafności interpretacji opartej na kontekście w scenariuszu Crohna oraz bardzo niską zgodność między ludzkimi anotatorami (**κ = 0,35**), co pokazuje, że sam problem jest subiektywny i niejednoznaczny [17].

Równoległy, lecz trochę inny trop rozwinęli **Ptaszynski, Dybala, Shi, Rzepka i Araki**. W ich ujęciu samo rozpoznanie emocji nie wystarcza, bo system powinien odróżnić ekspresję **adekwatną** od **nieadekwatnej** kontekstowo. Ich architektura łączy **tekstowy system analizy afektu** z **web miningiem / commonsense reasoning**, który szuka, jakie emocje są naturalnie skojarzone z danym kontekstem wypowiedzi. Na tej podstawie agent może zdecydować, czy powinien użytkownikowi współczuć, czy raczej pomóc mu regulować emocje. W tej pracy raportowano m.in. **72%** trafności web miningowego wydobywania skojarzeń emocjonalnych oraz – w zależności od rygoru oceny - około **45–50%** trafności dla oceny kontekstowej adekwatności emocji, a w warunkach łagodniejszych **80–85%** [10].

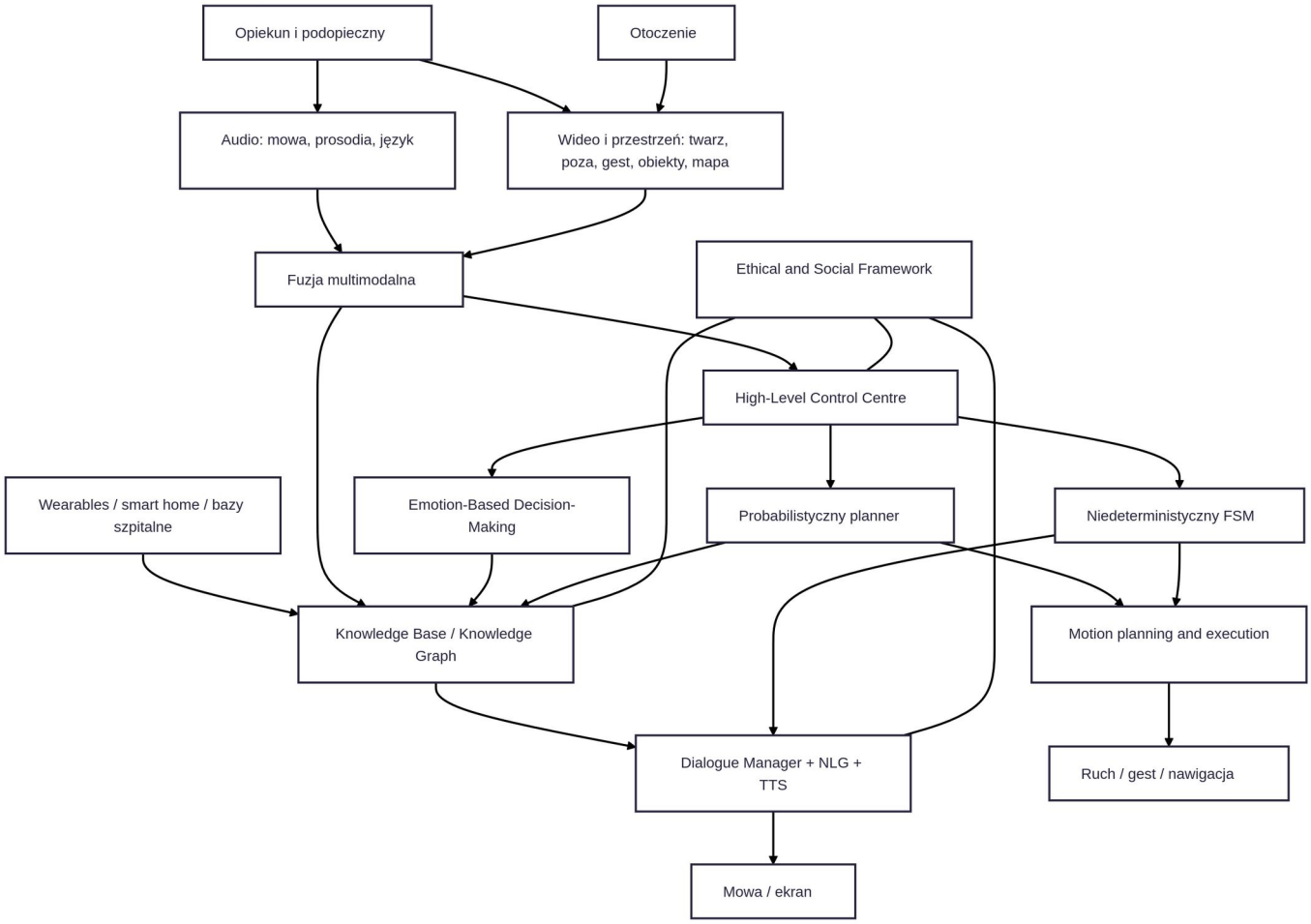
To ważny moment metodologiczny. Zarówno nurt Zhang, jak i nurt Ptaszyńskiego, przesuwają system afektywny z poziomu „mapowania sygnału na etykietę" ku poziomowi **interpretacji sytuacji, relacji i historii interakcji**. Nie są to jeszcze pełne modele przyczynowe w dzisiejszym sensie, ale są to podejścia **jawnie kontekstowe, sekwencyjne i semantyczne**, a więc bliższe rozumowaniu niż korelacji [18].

Porównanie nurtów poprzedzających CASIE

Nurt	Kontekst	Rozumowanie	Czasowość	Akcja	Niepewność	Miary ewaluacji
Zhang / e-drama / EMMA	temat dialogu, relacje między postaciami, profile emocjonalne, nastrój improwizacyjny, metafory	regułowo-semantyczne + modelowanie oparte o Markov chains / dynamic algorithm	tak, przez profile i historię dialogu	agent wybiera odpowiedzi wspierające przebieg improwizacji	szczątkowo ujęta przez confidence-driven response likelihood	Cohen's κ, accuracy ok. 68–69% dla modułu kontekstowego
Ptaszynski et al.	kontekst wypowiedzi + commonsense z Web mining	analiza afektu + ocena adekwatności emocji wobec kontekstu	umiarkowana	agent wybiera strategię: sympatyzowanie vs pomoc w regulacji emocji	brak formalnej kalibracji, ale jawna świadomość błędów kontekstu	balanced F-score, accuracy kontekstowej adekwatności, badania z agentami
CASIE 2021	audio, wideo, przestrzeń społeczna, mapa semantyczna, knowledge graph, dane z chmury, wearables, smart home	hybrydowe logiczno-probabilistyczne, high-level control, emotion-based decision-making, probabilistic planner	tak, osobne warstwy dla decyzji „tu i teraz" i długiego horyzontu	bardzo silne - pełny komponent planowania i wykonania ruchu/ dialogu	niepewność i niekompletność są uznane jako centralny problem, lecz bez pełnej procedury kalibracji	artykuł koncepcyjny; brak raportu pełnej ewaluacji end-to-end architektury

Architektura CASIE w robotyce zdrowotnej

Artykuł CASIE z 2021 roku definiuje architekturę, która **adaptuje** klasyczną pętlę robotyczną *sense–think–act* do interakcji społecznych wzbogaconych o przetwarzanie emocji. Autorzy podkreślają, że dane wejściowe mają służyć nie tylko bieżącej decyzji, lecz także budowaniu **pamięci długotrwałej** robota - tak, aby potrafił pamiętać twarze, tematy rozmów, obiekty w otoczeniu i kontekst sytuacyjny. [26]



W warstwie **percepcji językowo-głosowej** CASIE przewiduje identyfikację języka źródłowego, rozpoznanie emocji zgodnej z intencją dialogu, analizę prosodii oraz pipeline NLP typu hybrydowego - od tokenizacji i parsowania po NER i klasyfikację emocji. Autorzy proponują wykorzystanie zasobów takich jak **NRC EmoLex** i wcześniejszej platformy **SSIX**, a w części fuzji zakładają zarówno proste agregacje klasyfikatorów, jak majority voting, jak i bardziej zaawansowaną **feature-level fusion** z użyciem sieci głębokich. [27]

W warstwie **percepcji wizualno-przestrzennej** CASIE obejmuje równoległe pipeline'y rozpoznawania twarzy, pozy i mowy ciała, rozpoznawania obiektów, lokalizacji i mapowania. Szczególnie charakterystyczny element to **social map**: rozszerzenie mapy semantycznej i geometrycznej o lokalizację ludzi, ich pozę i dynamikę ruchu, z uwzględnieniem proxemiki, podejścia, prowadzenia i podążania. Tu już widać, że CASIE interesuje nie tylko „czy użytkownik jest smutny", ale też **jak ten stan emocjonalny i sytuacyjny wpływa na zachowania w przestrzeni społecznej**. [28]

Najciekawsza z punktu widzenia „rozumowania, a nie korelacji" jest jednak warstwa **High-Level Control Centre**. Składa się ona z trzech części:

1. pierwszej, **niedeterministycznego FSM** dla decyzji lokalnych i natychmiastowych;
2. drugiej, **Emotion-Based Decision-Making**, które rozstrzyga między potencjalnie sprzecznymi działaniami na podstawie stanów emocjonalnych wywnioskowanych z mowy, zdarzeń i obrazu;

3. trzeciej, **Robotic Planner**, czyli probabilistycznego planera budującego sekwencje działań o najwyższym prawdopodobieństwie osiągnięcia celu. Planner ma wprost radzić sobie z informacją **niekompletną, częściowo obserwowalną, stochastyczną i z niepewnymi post-conditionami**. [22]

Wiedza długotrwała robota jest ujęta jako **knowledge base / knowledge graph** z procesorem zapytań, optymalizatorem zapytań i cache'em. Baza wiedzy ma pełnić jednocześnie rolę pamięci długotrwałej, warstwy abstrakcji nad zewnętrznymi usługami i głównego zaplecza dla planowania oraz podejmowania decyzji. To ważne, bo oznacza, że CASIE nie traktuje emocji jako odrębnego „widgetu", lecz jako część szerszego stanu świata. [24]

Artykuł CASIE mocno akcentuje też warstwę **etyczno-społeczną**. Autorzy wymagają minimalizowania biasu na poziomie danych treningowych, modeli językowych i modeli rozpoznawania twarzy; podkreślają potrzebę transparentności i rozliczalności wobec potrzeb pacjenta, języka, płci, wieku i kontekstu opieki; oraz osadzają architekturę w unijnych i profesjonalnych ramach zaufanej AI. W przeciwieństwie do wielu prac z affect recognition, tutaj etyka nie jest dopiskiem, lecz osobnym komponentem architektury. [29]

Jednocześnie tekst wyraźnie ujawnia, że CASIE jest **programem badawczym**: autorzy omawiają dziewięć obszarów wyzwań, podają przewidywane komponenty i technologie, ale nie raportują pełnej walidacji całej zintegrowanej platformy na wdrożonym robocie. To istotne: CASIE jest architekturą ambitną i dojrzałą koncepcyjnie, ale nie jest jeszcze w tym artykule potwierdzonym, kompletnym benchmarkiem systemowym. [30]

CASIE wobec współczesnych modeli przyczynowych i niepewności

Najważniejsza granica między CASIE a dzisiejszymi **jawnie przyczynowymi** modelami emocji polega na tym, że CASIE mówi o **niekompletności, niepewności, częściowej obserwowalności i hybrydzie rozumowania logicznego z probabilistycznym**, ale nie formułuje jeszcze wprost **strukturalnego modelu przyczynowego** ani kontrfaktycznej procedury testowania „co by model przewidział, gdyby tło było inne". CASIE stawia to jako wyzwanie architektoniczne; nowsze prace zaczynają już dawać formalny aparat matematyczny. [22]

Dobrym kontrastem jest **CLEF** z CVPR 2024. Autorzy tej pracy pokazują, że w context-aware emotion recognition sam kontekst może być równie pomocny, co niebezpieczny, bo model uczy się **spurious shortcuts** między tłem a etykietą emocji. Proponują więc **uogólniony graf przyczynowy**, osobną gałąź modelującą bezpośredni niepożądany efekt kontekstu oraz procedurę porównywania wyników **faktycznych i kontrfaktycznych**, aby usunąć szkodliwy bezpośredni wpływ kontekstu z całkowitego efektu przyczynowego. To jest dokładnie ten poziom modelowania, którego CASIE jeszcze nie posiada, a który byłby naturalnym następcą jego ambicji. [16]

Podobnie rozwija się obszar **niepewności**. W rozmowach i interakcjach emocje są subiektywne, wieloznaczne i zmienne w czasie, czego symptomem były już stare wyniki Zhang: zgoda między ludzkimi anotatorami bywała niska, a sam agent miał porównywalne trudności interpretacyjne. Nowsze prace, takie jak **HU-Dialogue**, wprowadzają hierarchiczne modelowanie niepewności na

poziomie kontekstu i modalności, adaptując metody bayesowskie do ERC. Z kolei prace nad **conformal prediction** dla ERC pokazują, że można przejść od pojedynczej etykiety do **zbioru predykcji z gwarancją pokrycia** i redukować bias pokrycia między klasami emocji. [31]

To ważne, bo w ochronie zdrowia robot nie powinien być zmuszany do nadmiernej pewności. Jeśli rozpoznaje „lęk" tylko dlatego, że użytkownik siedzi w szpitalnej poczekalni, mamy klasyczny problem kontekstu jako pozornej przyczyny. Jeśli zaś jego posterior jest rozlany między „zmęczenie", „smutek" i „ból", poprawna odpowiedź systemu nie brzmi „smutek 0,81", lecz raczej: **„nie jestem dość pewny; czy chcesz powiedzieć, co czujesz?"** albo **„eskaluj do opiekuna"**. Tę logikę wspierają zarówno współczesne prace o niepewności w ERC, jak i unijne wymogi dotyczące nadzoru człowieka, przejrzystości i bezpieczeństwa technicznego. [32]

Warto tu dodać ostrożność płynącą z badań HRI. Przegląd „Four Critical Lessons" podkreśla, że w interakcjach sytuowanych **sześć podstawowych kategorii emocji nie zawsze jest odpowiednie**, sama poprawa accuracy nie jest dobrym celem, a modele afektu **słabo generalizują między kontekstami**. To bardzo mocna krytyka dla każdej wdrożeniowej wersji CASIE, jeśli nie będzie ona uczona i oceniana w konkretnym środowisku opiekuńczym. [33]

Tabela porównawcza CASIE z nowocześniejszymi rozszerzeniami

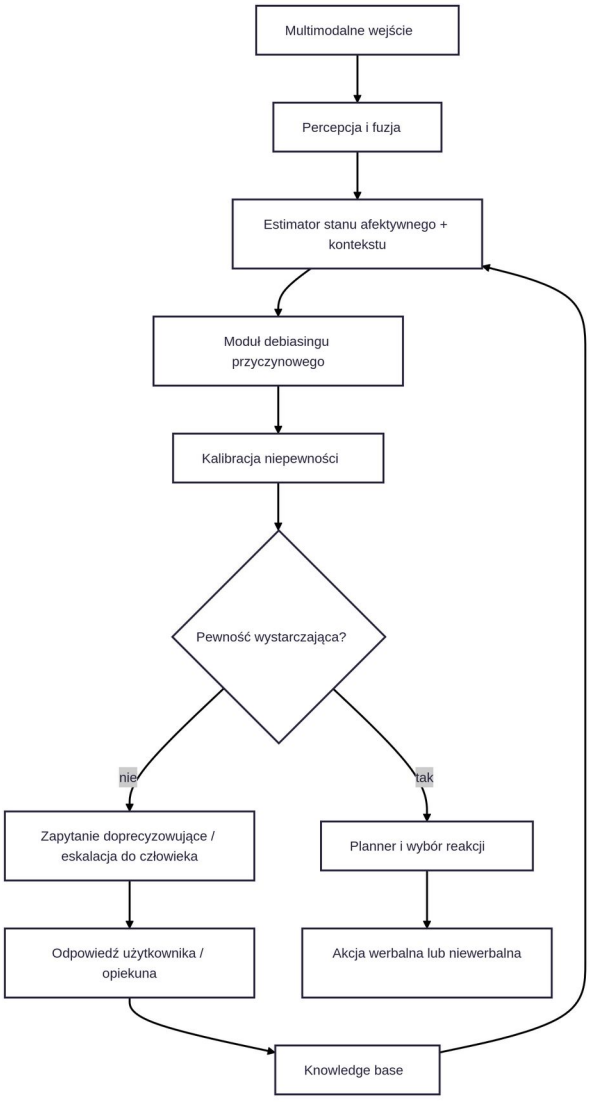
Podejście	Zdolność przyczynowa	Niepewność	Możliwość abstencji / dopytania	Interpretowalność	Wykonalność czasu rzeczywistego	Główna słabość
Zhang / EMMA	niska do umiarkowanej; kontekst i sekwencja, ale bez jawnego grafu przyczynowego	heurystyczna pewność odpowiedzi	pośrednio możliwa, lecz nieformalna	umiarkowana, bo oparte na profilach kontekstu i regułach	dobra dla dialogu tekstowego	mała skala i domenowość scenariuszy
Ptaszynski et al.	umiarkowana; ocena adekwatności emocji wobec kontekstu, ale bez formalnego SCM	brak nowoczesnej kalibracji	agent wybiera strategię reakcji, lecz bez abstain policy	wysoka konceptualnie	umiarkowana	zależność od jakości commonsense/web mining
CASIE 2021	umiarkowana; hybryda logiczno-probabilistyczna i planner, lecz bez jawnych kontrfaktyków	rozpoznana jako centralna potrzeba, ale nieoperacjonalizowana kalibracyjnie	nieopisana explicite w architekturze	wysoka na poziomie modułów i wiedzy	projektowana pod real-time i edge/cloud, ale kosztowna obliczeniowo	brak pełnej walidacji end-to-end w publikacji
CLEF 2024	wysoka; jawny graf przyczynowy i wnioskowanie kontrfaktyczne	pośrednia, przez oddzielenie użytecznego efektu kontekstu od biasu	nie jest to główny cel pracy	umiarkowana	potencjalnie dobra jako nakładka model-agnostic	skupia się na debiasingu, nie na całym HRI/planning stacku
HU-Dialogue + conformal ERC	niska przyczynowo, wysoka w niepewności	wysoka; uncertainty hierarchiczne i zbiory predykcji z coverage guarantees	wysoka; naturalna baza dla „abstain / ask clarification"	umiarkowana	dobra dla systemów konwersacyjnych	nie rozwiązuje sama z siebie problemu przyczynowości ani planowania

Ryzyka, regulacje i rekomendacje dla nowoczesnego CASIE

W świetle prawa unijnego emocjonalne systemy robotyczne działają dziś w dużo bardziej wymagającym otoczeniu niż w czasie formułowania wcześniejszych frameworków. **AI Act** definiuje „emotion recognition system" jako system AI służący do identyfikowania lub wnioskowania o emocjach lub intencjach osób fizycznych na podstawie danych biometrycznych. W unijnym podsumowaniu po polsku wskazano też zakaz wnioskowania o emocjach w **miejscach pracy** i **instytucjach edukacyjnych**, z wyjątkami m.in. dla celów medycznych lub bezpieczeństwa. To nie eliminuje CASIE dla opieki zdrowotnej, ale wyraźnie podnosi wymogi dotyczące uzasadnienia celu, proporcjonalności, dokumentacji i nadzoru człowieka. [39]

Równolegle europejskie wytyczne dla godnej zaufania AI oraz **ALTAI** wymagają: przewodniej i nadzorczej roli człowieka, technicznej solidności i bezpieczeństwa, ochrony prywatności i danych, przejrzystości, niedyskryminacji, dobrostanu społecznego oraz rozliczalności. To niezwykle dobrze współgra z etycznym komponentem CASIE, ale równocześnie odsłania jego dzisiejszy brak: aby spełnić te wymogi praktycznie, system powinien umieć **wycofać się z decyzji, gdy jest niepewny**, a nie tylko wyprodukować najbardziej prawdopodobną etykietę emocji. [40]

Warto też studzić oczekiwania wdrożeniowe. Przegląd parasolowy nad społecznie wspierającymi robotami w zdrowiu i opiece pokazał, że choć liczne badania sygnalizują poprawę nastroju, interakcji społecznych, samotności, stresu czy bólu, to łączna meta-analityczna ocena nie wykazała jednoznacznego efektu dla jakości życia, lęku i depresji. Innymi słowy: wartość robotów społecznych istnieje, ale **nie należy jej automatycznie utożsamiać z niezawodnym rozpoznawaniem emocji**.



Dlatego najbardziej sensowna ścieżka rozwojowa to nie „wdrożyć CASIE tak, jak opisano w 2021", lecz zbudować **wariant z kontrolowaną niepewnością i aktywnym doprecyzowaniem**. Taki system powinien łączyć:

1. interpretowalną warstwę wiedzy i planowania z CASIE,
2. kontrfaktyczne odszumianie kontekstu w stylu CLEF,
3. warstwę niepewności typu HU-Dialogue / conformal prediction dla decyzji o abstencji.

W praktyce architektura typu **CASIE-2** powinna działać według prostej zasady: jeśli kontekst i sygnały są spójne, system planuje działanie; jeśli są słabe, sprzeczne lub obciążone biasem kontekstowym, system **zadaje pytanie doprecyzowujące**, przełącza się na reakcję neutralno-

wspierającą albo eskaluje do personelu. Takie zachowanie jest bardziej zgodne zarówno z psychologiczną naturą emocji, jak i z wymaganiami zaufanej AI.

Werdykt badawczy

Jeśli szukać **najbliższego literaturowego odpowiednika „CASIE framework”**, to właściwą odpowiedzią jest **CASIE 2021 jako referencyjna architektura robotyki afektywnej dla zdrowia**, osadzona w starszej genealogii **context-sensitive affect sensing** i **context-aware emotional intelligence**. Jeśli natomiast pytanie dotyczy systemów, które rzeczywiście „rozumują przyczynowo i kontekstowo”, to CASIE trzeba dziś czytać jako **bardzo cenną bazę systemową**, ale wymagającą uzupełnienia o współczesne komponenty: **jawne modele przyczynowe, formalne modelowanie niepewności, abstencję i aktywne dopytywanie**.

Największa luka w literaturze nie dotyczy już samej percepcji afektu, lecz **zintegrowania** pięciu rzeczy naraz: bogatego kontekstu społecznego, rozumowania przyczynowego, belief-state pod częściową obserwowalnością, kalibracji niepewności oraz polityki bezpiecznego działania. CASIE wykonało ważny krok w stronę takiej integracji, ale nie zamknęło tego programu badawczego. Właśnie dlatego pozostaje ono wartościowe: nie jako ostateczne rozwiązanie, lecz jako **ambitny szkielet architektury, który nadal wyznacza sensowny kierunek badań**. [45]

Literatura

1. Vasiliu, L., Cortis, K., McDermott, R., Kerr, A., Peters, A., Hesse, M., Hagemeyer, J., Belpaeme, T., McDonald, J., Villing, R., Mileo, A., Caputo, A., Scriney, M., Griffiths, S., Koumpis, A., & Davis, B. (2021). CASIE – Computing affect and social intelligence for healthcare in an ethical and trustworthy manner. *Paladyn, Journal of Behavioral Robotics, 12*(1), 437–453. https://doi.org/10.1515/pjbr-2021-0026
2. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R., & Araki, K. (2009). *Towards context aware emotional intelligence in machines: Computing contextual appropriateness of affective states*. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)* (pp. 1172–1177). IJCAI.
3. Zhang, L., Barnden, J. A., Hendley, R. J., & Wallington, A. M. (2006). *Developments in affect detection in e-drama*. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 17–20). Association for Computational Linguistics.
4. Zhang, L. (2010). Metaphor interpretation and context-based affect detection. In J. Huang & D. Jurafsky (Eds.), *COLING 2010: Posters* (pp. 1523–1531). Association for Computational Linguistics.
5. Zhang, L. (2011). Exploration on context-sensitive affect sensing in an intelligent agent. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, & K. R. Thórisson (Eds.), *Intelligent Virtual Agents: 11th International Conference, IVA 2011* (Lecture Notes in Computer Science, Vol. 6895, pp. 427–430). Springer. https://doi.org/10.1007/978-3-642-23974-8_44
6. Zhang, L., & Barnden, J. A. (2011). Context-sensitive affect sensing and metaphor identification in virtual drama. In *Affective Computing and Intelligent Interaction: 4th International Conference,*

ACII 2011 (Lecture Notes in Computer Science, Vol. 6974, pp. 322–331). Springer. https://doi.org/10.1007/978-3-642-24571-8_18

7. Zhang, L., & Hossain, A. (2012). Contextual affect sensing and metaphor interpretation. *Entertainment Computing, 3*(3), 81–88. https://doi.org/10.1016/j.entcom.2011.12.002
8. Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research, 5*(4), 269–306. https://doi.org/10.1016/j.cogsys.2004.02.002
9. Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys, 3*, 96–146. https://doi.org/10.1214/09-SS057
10. Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.
11. Ha, D., & Schmidhuber, J. (2018). *World models*. arXiv. https://doi.org/10.48550/arXiv.1803.10122
12. Yang, D., Yang, K., Li, M., Wang, S., Wang, S., & Zhang, L. (2024). Robust emotion recognition in context debiasing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
13. Roohi, S., Skarbez, R., & Nguyen, H. D. (2024). Reliable uncertainty estimation in emotion recognition in conversation using conformal prediction framework. *Natural Language Processing*. Advance online publication. https://doi.org/10.1017/nlp.2024.57
14. Gunes, H., & Churamani, N. (2023). Affective computing for human-robot interaction research: Four critical lessons for the hitchhiker. In *Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023)*.
15. Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, A. (2017). Emotion recognition in context. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)* (pp. 1960–1968). IEEE. https://doi.org/10.1109/CVPR.2017.212