

přerušení nebo pokračování medikace a screeningu laboratorních výsledků), tvorbou edukačních materiálů pro pacienty nebo orientací v ověřených informačních zdrojích a doporučených postupech [8, 9, 10]. Jde však o úlohy podpůrné povahy a jejich přínos je zatím doložen převážně srovnávacími studiemi a studiemi proveditelnosti, nikoli prospektivními daty o dopadu na klinicky relevantní výsledky – například hodnocením čitelnosti předoperačních instrukcí nebo kvality odpovědí na dotazy pacientů [8]. Prospektivní data o dopadu na klinicky relevantní výsledky zatím chybí. Technickou dostupnost obecných jazykových modelů proto nelze zaměňovat za doložený klinický přínos. Jde o prostředky použitelné pro jasně vymezené, lékařem ověřované podpůrné úkoly, nikoli o klinicky validované nástroje pro konkrétní perioperační indikaci [8, 9]. Systémy ambientní klinické dokumentace mohou převádět mluvený rozhovor do strukturovaného záznamu.

WHO uvádí, že velké jazykové modely mohou snižovat administrativní zátěž lékařů spojenou s vedením zdravotnické dokumentace, zároveň však varuje, že tyto funkce nemají být plně automatizovány a vyžadují lidskou kontrolu [9]. Výstupy těchto systémů musí být lékařem vždy ověřeny z hlediska správnosti, úplnosti, klinického kontextu, možných halucinací, zkreslení (bias) a ochrany osobních údajů [8, 9]. Jazyková přesvědčivost výstupu LLM a jeho faktická správnost jsou dvě různé věci. Právě jejich záměna představuje při klinickém použití největší riziko.

Další oblastí, kde se AI nástroje mohou v této fázi péče objevit, je predikce rizika. Přínosné však mohou být jen tehdy, pokud nepřidávají jen další „rizikové skóre“, ale pomáhají rozhodnout, který pacient vyžaduje změnu perioperačního plánu ideálně ještě před vstupem na operační sál. Výstup musí mít konkrétní důsledek, např. doplnění vyšetření, konzilium, úpravu medikace, změnu monitorace, rezervaci pooperačního lůžka nebo cílenější rozhovor o riziku [2, 3].

Typickým příkladem je predikce pooperační mortality nebo závažných komplikací u vysoce rizikových či akutně přidaných výkonů do operačního programu. Model zde nemá automatizovat rozhodnutí o tom, zda má být operace provedena, ale upozornit na pacienta vhodného k rozšířenému anesteziologickému zhodnocení. Klinický přínos závisí nejen na schopnosti modelu správně identifikovat rizikové pacienty, ale i na tom, zda je výstup kalibrován pro lokální populaci, srozumitelně prezentován a navázán na konkrétní rozhodovací kroky [3, 5].

Jako konkrétní příklad lze uvést MySurgeryRisk, algoritmus vyvinutý a validovaný v monocentrické kohortě 51 457 pacientů podstupujících velký chirurgický výkon. Model využíval předoperačně dostupná data z elektronické zdravotnické dokumentace (EHR) k odhadu rizika osmi závažných pooperačních komplikací a mortality. Hodnoty AUC (area under the receiver operating characteristic curve, plocha pod ROC křivkou) dosahovaly 0,82–0,94 pro komplikace a 0,77–0,83 pro mortalitu v různých časových horizontech [11]. Při interpretaci těchto výsledků je nutné zohlednit interní povahu vývoje a validace modelu na datech z elektronické zdravotnické dokumentace jednoho centra a skutečnost, že sami autoři upozorňují na potřebu dalšího ověření proveditelnosti implementace v reálném klinickém provozu. Studie tak mimo jiné dokládá, že vysoká predikční přesnost v interní validaci sama o sobě

nezajišťuje bezpečné, přenositelné ani klinicky přínosné použití modelu v odlišném lokálním prostředí.

Novější retrospektivní studie na souboru 111 888 výkonů z jednoho centra navíc oddělila předoperační a intraoperační data a zjistila, že modely využívající výhradně předoperační údaje dosahovaly při predikci pooperačních komplikací jen mírně nižší diskriminační schopnosti než modely kombinující obě fáze (rozdíl AUC 0,002–0,032). Tato práce však zůstala na úrovni interní validace bez externího ověření [12].

Otázku, zda taková predikce skutečně zlepší rozhodování anesteziologa, zkoumala monocentrická prospektivní randomizovaná studie Perioperative ORACLE. V ní 89 anesteziologů hodnotilo riziko pooperační mortality a akutního poškození ledvin u 5 071 případů buď s přístupem k predikcím strojového učení, nebo bez něj. Prospektivní výkon samotných modelů byl nižší než při retrospektivní validaci (AUC 0,81 pro mortalitu a 0,77 pro akutní poškození ledvin). Přístup k predikcím přitom nevedl ke statisticky významnému zlepšení diskriminační schopnosti rizikového odhadu kliniků (AUC pro mortalitu 0,793 vs. 0,780; pro akutní poškození ledvin 0,734 vs. 0,688). Predikce kliniků se sice v asistované skupině častěji shodovaly s modelem, tato shoda však nastávala i tehdy, když byl výstup modelu nesprávný [13]. Studie ukazuje, že ani technicky funkční predikční model nemusí sám o sobě zlepšit rizikový odhad kliniků, není-li jeho výstup spolehlivý a dobře integrovaný.

Podobně lze chápat predikci pooperačního deliria. U starších pacientů neslouží AI k „diagnóze deliria před výkonem“, ale k odhadu individuálního rizika a k podpoře plánování perioperační péče a cílenějšího pooperačního sledování. Ve studii BioCog, prospektivní observační kohortě ze dvou terciárních center, se u pacientů ve věku ≥ 65 let bez preexistující demence pooperační delirium vyskytlo přibližně ve 20 %. Nejvýkonnější interně validovaný model kombinující předoperační klinická data, charakteristiky výkonu a perioperační laboratorní změny dosáhl AUC 0,83, zatímco model využívající pouze předoperační klinická data dosáhl AUC 0,76 [14]. Při interpretaci je nutné zohlednit interní povahu validace bez nezávislého externího ověření, použití části nerutinních vstupních dat a dosud neprokázaný dopad na klinické rozhodování nebo výskyt deliria [14]. Zjištění, zda identifikace rizikových pacientů tímto modelem skutečně vede ke snížení incidence pooperačního deliria nebo ke zlepšení klinicky relevantních výsledků, zůstává nedostatečně doloženo.

Obdobně koncipovanou oblast představuje AI analýza předoperačního EKG. Retrospektivní studie ukazují, že hluboké učení z předoperačního dvanáctivodového EKG může zpřesnit odhad mortality a závažných kardiovaskulárních příhod oproti běžným skóre. Přes externí validaci jednoho z modelů však chybí prospektivní průkaz klinického přínosu; oblast proto nelze považovat za rutinně implementovanou [15, 16]. Další předoperační nástroje se zabývají automatickou klasifikací fyzického stavu podle ASA, predikcí potřeby pooperačního přijetí na jednotku intenzivní péče a predikcí obtížného zajištění dýchacích cest. Přehled Dost et al. uvádí tyto aplikace jako příklady slibných podpůrných nástrojů. Publikované práce popisují povzbudivou shodu s hodnocením anesteziologa nebo dobrou diskriminační schopnost, zatímco přínos pro klinicky relevantní výsledky a proveditelnost rutinního nasazení zůstávají nedostatečně ověřeny [2].