Ich möchte in meiner Arbeit diese Algorithmen damit vergleichen, wie KI, Antworten auf diese Fragetypen bewerten würde. Anhand von den Kriterien, die eine Antwort als korrekt oder inkorrekt klassifizieren, basierend auf Usereinstellungen (Groß/Kleinschreibung beachten, bei Schätzfragen % angeben, Rechtschreibfehler bis zu einem bestimmten level zulassen und wenig bis keine Punkte abziehen, Synonyme akzeptieren) und natürlich auf Genauigkeit, Effizienz und Kosten.

Algorithmen:

Multiple Choice / Entweder-Oder:

Vergleichsbasierte Algorithmen (Vergleichsmatrix): Sehr einfach zu implementieren.

Gewichtete Punktesysteme: Ebenfalls leicht mit normaler Logik (z. B. JSON-Strukturen) umzusetzen.

Ein-Wort-Antworten:

String-Matching-Algorithmen: Leicht (z. B. lower().strip() + Vergleich).

Stemming und Lemmatisierung: Mit z. B. NLTK oder spaCy in Python gut machbar.

Phonetische Ähnlichkeit: Mit fuzzy oder Metaphone realisierbar – einfacher als semantische Methoden.

Rechenfragen:

Numerische Berechnung: Simpler Vergleich mit Toleranzbereich – sehr einfach.

Schätzfragen:

Toleranzbereiche: Leicht mit einfacher Differenzprüfung (z. B. abs(user - correct) < tol).

Lückentext:

Wortübereinstimmung: Direkt umsetzbar über Vergleich pro Lücke.

Freitext:

Keyword-Matching: Mit Set- oder Listenvergleichen einfach machbar.

N-gram-Matching: Einfach mit Tokenisierung und Sliding Window.

Ich würde die fragetypen so einstufen:  
  
Bei Multiple Choice und Entweder-Oder sind herkömmliche Algorithmen besser, da zuvorbestimmte lösung und algorithmen einfacher und effizienter

Ein-Wort-Antworten, Rechenfragen und Schätzfragen sollte KI schon in erwegung bezogen werden, da sie Synonyme besser erkennt, bei Rechenfragen, wenn der Rechenweg mitggeben ist noch teilpunkte geben kann, wenn die lösung falsch ist, und bei Schätzfragen kann sie, wenn ein Toleranzbereich in % mitgegeben wird, die Antwort individuell und passend bewerten.   
Da sind herkömmliche alg. Sehr viel eingeschränkter und unpräziser

Bei Lückentexten und Freitexten ist KI sowieso besser da sie natürliche Sprache semantisch und kontextbasiert viel besser bewerten kann. Sie kann synonyme, rechtschreibfehler und groß und kleinschreibung erkennen und trotzdem (mit punktabzug) fair bewerten. Alg. Tun sich da besonders schwer und sind einfach schlechetr als KI