



David Gewirtz/Zdnet

Deepseek explodiu na consciência do mundo no fim de semana passado. Se destaca por três razões poderosas:

1. É um chatbot da AI da China, em vez dos EUA
2. É de código aberto.
3. Ele usa muito menos infraestrutura do que as grandes ferramentas de IA que estamos olhando.

Além disso: os pesquisadores da Apple revelam o molho secreto por trás do Deepseek AI

Dadas as preocupações do governo dos EUA sobre o Tiktok e o possível envolvimento do governo chinês nesse código, uma nova IA emergente da China deve gerar atenção. Radhika Rajkumar, da ZDNET, mergulhou profundamente nesses problemas em seu artigo por que o Deepseek da China poderia estourar nossa bolha de IA.

Neste artigo, estamos evitando a política. Em vez disso, estou colocando o Deepseek V3 e o Deekseek R1 no mesmo conjunto de testes de codificação de IA que joguei em 10 outros modelos de idiomas grandes. De acordo com o próprio DeepSeek:

- **Escolha v3** Para tarefas que requerem profundidade e precisão (por exemplo,



resolvendo problemas de matemática avançados, gerando código complexo).

- **Escolha R1** Para aplicativos sensíveis à latência e de alto volume (por exemplo, automação de suporte ao cliente, processamento básico de texto).

Você pode escolher entre R1 e V3 clicando no botão pequeno na interface de bate -papo. Se o botão estiver azul, você está usando R1.





A resposta curta é a seguinte: impressionante, mas claramente não perfeita. Vamos cavar.

O que você vai ler:



- [Teste 1: Escrevendo um plugin do WordPress](#)
- [Teste 2: Reescrevendo uma função de string](#)
- [Teste 3: Encontrar um bug irritante](#)
- [Teste 4: Escrevendo um script](#)
- [Pensamentos finais](#)

Teste 1: Escrevendo um plugin do WordPress

Este teste foi na verdade o meu primeiro teste das proezas de programação do ChatGPT, naquela época. Minha esposa precisava de um plug -in para o [WordPress](#) que a ajudaria a executar um dispositivo de envolvimento para seu grupo on -line.

Além disso: a melhor IA para codificar em 2025 (e o que não usar)

Suas necessidades eram bastante simples. Precisava obter uma lista de nomes, um nome por linha. Ele então teve que classificar os nomes e, se houvesse nomes duplicados, separasse-os para que eles não estivessem listados lado a lado.

Eu realmente não tive tempo de codificá -lo para ela, então decidi dar à IA o desafio por um capricho. Para minha enorme surpresa, funcionou.

Desde então, tem sido o meu primeiro teste para a AIS ao avaliar suas habilidades de programação. Requer que a IA saiba como configurar o código para a estrutura do WordPress e siga solicitações com clareza o suficiente para criar a interface do usuário e a lógica do programa.



Captura de tela de David Gewirtz/Zdnet

Apenas cerca de metade do AIS que testei pode passar completamente neste teste. Agora, no entanto, podemos adicionar mais um ao círculo do vencedor.



Captura de tela de David Gewirtz/Zdnet

O Deepseek V3 criou a interface do usuário e a lógica do programa exatamente como especificado. Quanto ao Deepseek R1, bem, esse é um caso interessante. O aspecto “raciocínio” do R1 fez com que a IA cuspir 4502 palavras de análise antes de compartilhar o código.



Captura de tela de David Gewirtz/Zdnet

A interface do usuário parecia diferente, com áreas de entrada muito mais amplas. No entanto, tanto a interface do usuário quanto a lógica funcionaram, então o R1 também passa neste teste.



Captura de tela de David Gewirtz/Zdnet

Até agora, o Deepseek V3 e o R1 passaram um dos quatro testes.

Teste 2: Reescrevendo uma função de string

Um usuário reclamou que não conseguiu inserir dólares e centavos em um campo de entrada de doações. Como escrito, meu código só permitia dólares. Então, o teste envolve dar à IA a rotina que eu escrevi e pedindo para reescrever para permitir dólares e centavos

Além disso: meu recurso de chatgpt favorito acabou de ficar muito mais poderoso



Geralmente, isso resulta na IA gerando algum código de validação de expressão regular. A Deepseek gerou código que funciona, embora haja espaço para melhorias. O código que o Deepseek V2 escreveu foi desnecessariamente longo e repetitivo, enquanto o raciocínio antes de gerar o código em R1 também foi muito longo.

Minha maior preocupação é que ambos os modelos da validação Deepseek garantem a validação de até 2 casas decimais, mas se um número muito grande for inserido (como 0,30000000000000004), o uso do Parsefloat não tem conhecimento de arredondamento explícito. O modelo R1 também usou a conversão de números do JavaScript sem verificar as entradas de casos de borda. Se os dados ruins voltarem de uma parte anterior da expressão regular ou de uma não coragem, o que chegar a essa conversão, o código falharia.

É estranho, porque o R1 apresentou uma lista muito agradável de testes para validar contra:



Captura de tela de David Gewirtz/Zdnet

Então, aqui, temos uma decisão dividida. Estou dando o objetivo do DeepSeek V3 porque nenhum desses problemas que seu código produziu fará com que o programa quebrasse quando executado por um usuário e geraria os resultados esperados. Por outro lado, tenho que dar uma falha em R1, porque se algo não for uma string de alguma forma entrar na função numérica, uma falha ocorrerá.

E isso dá ao Deepseek V3 duas vitórias em quatro, mas o Deepseek R1 apenas uma vitória em quatro até agora.



Teste 3: Encontrar um bug irritante

Este é um teste criado quando tive um bug muito irritante que tive dificuldade em rastrear. Mais uma vez, decidi ver se o ChatGPT poderia lidar com isso, o que aconteceu.

O desafio é que a resposta não é óbvia. Na verdade, o desafio é que há uma resposta óbvia, com base na mensagem de erro. Mas a resposta óbvia é a resposta errada. Isso não apenas me pegou, mas regularmente pega algumas das IAs.

Também: ChatGpt Plus ou Pro Walth a vale a pena? Veja como eles se comparam à versão gratuita

Resolver esse bug requer entender como as chamadas específicas da API no WordPress funcionam, poder ver além da mensagem de erro no próprio código e depois saber onde encontrar o bug.

O Deepseek V3 e o R1 passaram este com respostas quase idênticas, levando -nos a três de quatro vitórias para V3 e duas em quatro vitórias por R1. Isso já coloca seek à frente de Gêmeos, Copilot, Claude e Meta.

Deepseek marcará um home run para a V3? Vamos descobrir.

Teste 4: Escrevendo um script

E outro morde a poeira. Este é um teste desafiador, pois exige que a IA compreenda a interação entre três ambientes: AppleScript, o modelo de objeto Chrome e uma ferramenta de script Mac chamada Maestro do Teclado.

Eu teria chamado isso de teste injusto porque o teclado Maestro não é uma ferramenta de programação convencional. Mas o ChatGPT lidou com o teste facilmente, entendendo exatamente qual parte do problema é tratada por cada ferramenta.

Além disso: como o chatgpt digitalizou 170k linhas de código em segundos, economizando horas de trabalho

Infelizmente, nem o Deepseek V3 ou o R1 tinha esse nível de conhecimento. Nenhum dos modelos sabia que precisava dividir a tarefa entre as instruções para o teclado Maestro e o Chrome. Ele também tinha um conhecimento bastante fraco do AppleScript, escrevendo rotinas personalizadas para AppleScript, nativas do idioma.

Estranhamente, o modelo R1 também falhou porque fez um monte de suposições incorretas. Ele assumiu que sempre existe uma janela da frente, o que definitivamente não é o caso. Ele também assumiu que o programa de corrida frontal atualmente sempre seria o Chrome, em vez de verificar explicitamente para ver se o Chrome estava em execução.



Isso deixa o Deepseek V3 com três testes corretos e uma falha e o Deepseek R1 com dois testes corretos e duas falhas.

Pensamentos finais

Descobri que a insistência de Deepseek em usar um endereço de email em [nuvem](#) pública como Gmail.com (em vez do meu endereço de e-mail normal com meu domínio corporativo) era irritante. Ele também teve uma série de falhas de capacidade de resposta que fizeram fazer esses testes mais tempo do que eu gostaria.

Também: Como usar o chatgpt para escrever código: o que faz bem e o que não

Eu não tinha certeza de que seria capaz de escrever este artigo porque, durante a maior parte do dia, recebi esse erro ao tentar me inscrever:

Os serviços on-line da Deepseek enfrentaram recentemente ataques maliciosos em larga escala. Para garantir o serviço contínuo, o registro é temporariamente limitado a +86 números de telefone. Os usuários existentes podem fazer login como de costume. Obrigado pela sua compreensão e apoio.

Então, entrei e consegui executar os testes.

O Deepseek parece ser excessivamente loquaz em termos do código que ele gera. O código AppleScript no Teste 4 estava errado e excessivamente longo. O código de expressão regular no Teste 2 estava correto no V3, mas poderia ter sido escrito de uma maneira que o tornou muito mais sustentável. Falhou em R1.

Também: Se o ChatGPT produz código gerado pela IA para o seu aplicativo, a quem realmente pertence?

Definitivamente, estou impressionado que o Deepseek V3 venceu Gêmeos, Copilot e Meta. Mas parece estar no nível antigo do GPT-3.5, o que significa que definitivamente há espaço para melhorias. Fiquei decepcionado com os resultados do modelo R1. Dada a escolha, eu ainda escolheria o ChatGPT como meu ajudante de código de programação.

Dito isto, para uma ferramenta nova em folha em infraestrutura muito mais baixa do que as outras ferramentas, isso pode ser uma IA para assistir.

O que você acha? Você já tentou Deepseek? Você está usando algum AIS para suporte a programação? Informe -nos nos comentários abaixo.



Você pode seguir minhas atualizações do projeto diário nas mídias sociais. Certifique -se de assinar meu boletim informativo semanal de atualização e siga -me no [Twitter](#)/x em [@Davidgewirtz](#) no Facebook em facebook.com/davidgewirtz, no Instagram em Instagram.com/davidgewirtz, em bluesky em [@Davidgewirtz.com](#) e no YouTube em YouTube.com/davidgewirtz.