

El impacto de asistentes basados en IA en la enseñanza-aprendizaje de la programación

Francisco de Sande

Departamento de Ingeniería Informática y de Sistemas

Universidad de La Laguna

38200 La Laguna. S/C de Tenerife

fsande@ull.es

Resumen

Abstract

Abstract

Abstract

Palabras clave

Programación IA ChatGPT GitHub Copilot Enseñanza Informática Evaluación

1. Introducción

La programación es una actividad transversal a cualquier rama de la informática y su importancia es compartida por cualquier especialización en esta titulación. *Informática Básica* (IB, de ahora en adelante) es la primera asignatura en la que el alumnado toma contacto con esta materia y en la que ha de aprender los fundamentos de la misma. Los conceptos objeto de estudio son comunes, con pequeñas variaciones, a cualquier lenguaje de programación orientado a objetos, que es el paradigma objeto de estudio en IB.

IB es una asignatura de 6 créditos que se imparte en el primer cuatrimestre del primer curso del Grado en Ingeniería Informática en la Escuela Superior de Ingeniería y Tecnología de la Universidad de La Laguna. Se trata de la primera asignatura (y única en ese primer cuatrimestre) de perfil eminentemente informático que cursa el alumnado del título de Grado.

El número de estudiantes que habitualmente cursa la asignatura está en torno a los 250. Los contenidos de la asignatura pueden consultarse en la Guía Docente ¹. de la misma. Al margen de tres temas dedicados a

una introducción a las Bases de Datos, Redes y Sistemas Operativos, el grueso de los contenidos (en torno a 12 de las 15 semanas del cuatrimestre) se dedican a introducir al alumnado en la materia de Programación, siendo C++ el lenguaje vehicular elegido para estudiar la materia. Se trata de una asignatura con una importante proporción de contenidos prácticos en la que cada estudiante recibe 4 horas presenciales de clase a la semana distribuidas del siguiente modo:

- 2 horas dedicadas al estudio de contenidos teóricos
- 1 hora dedicada a la resolución de Problemas
- 1 horas de Prácticas

Se describe a continuación el tipo de actividades que se desarrolla en cada una de este tipo de sesiones.

Las sesiones destinadas a contenidos teóricos se imparten con el formato de clase expositiva y se dedican, con el uso de transparencias que el alumnado tiene disponibles a través del aula virtual, al estudio de los contenidos de la asignatura. Junto a las transparencias, el alumnado dispone de una serie de pequeños programas que sirven de ilustración a los contenidos que se estudian en clase. Se recomienda al alumnado estudiar esos programas para afianzar sus conocimientos de cada tema.

En las sesiones de problemas el profesorado utiliza un terminal cuya pantalla se proyecta a toda la clase para resolver algunos problemas seleccionados, resolviendo al mismo tiempo las dudas que puedan surgir en esa resolución.

La Guía docente establece que por cada hora de trabajo presencial, cada estudiante debería dedicar en promedio 1,5 horas de trabajo autónomo, de modo que se espera unas 6 horas de trabajo autónomo semanal por parte de cada estudiante. A pesar de que el peso de los contenidos prácticos de la asignatura suponen solo un 20 % del total de la calificación de la asignatura, se espera que la mayor parte de ese tiempo se destine al diseño y desarrollo de programas de complejidad creciente conforme avanza el desarrollo del curso, y que se evalúan en las sesiones prácticas.

¹Guía Docente de IB

https://www.ull.es/apps/guias/guias/view_guide/34182/

Cada semana al alumnado se le propone la realización de una práctica, consistente en un cierto número (cinco es un número habitual) de programas en C++ relacionados con algún tema estudiado. Esos programas tienen el propósito de servir de “entrenamiento” para que el alumnado afiance conocimientos a lo largo de la semana de la que disponen para realizar esos ejercicios. La sesión semanal de prácticas se destina a la evaluación de esos conocimientos a través de la realización de ejercicios de programación de complejidad similar a los que han sido propuestos con antelación. En las últimas prácticas de la asignatura los ejercicios de evaluación suelen ser simples modificaciones de los que se han propuesto para realizar con antelación.

2. Motivación

Más allá del ineludible aprendizaje de los conceptos básicos de la materia, es un hecho bien conocido que la práctica es fundamental para aprender a programar ordenadores. Habitualmente el profesorado asigna problemas de programación al alumnado para ayudarles a adquirir esta destreza, y el incremento de las habilidades de una programadora pasa ineludiblemente por muchas horas de dedicación a la realización de programas de complejidad creciente. Forzando un poco las similitudes podríamos decir que la programación se asemeja a las habilidades para conducir un automóvil: cualquier persona con permiso para conducir puede decir que conduce correctamente, pero acreditar que se es una buena conductora requiere muchas horas de práctica y exposición a situaciones infrecuentes. Del mismo modo cualquier informático dirá que sabe programar pero sus habilidades en esta materia dependerán muy directamente de las horas de práctica que haya dedicado a la misma. Siguiendo esta idea, recomendamos encarecidamente al alumnado de IB que realice cuantos ejercicios prácticos sean capaces para, de forma progresiva, ir incrementando sus capacidades como programadoras.

Las prácticas de la asignatura se convierten pues en una oportunidad para que el alumnado mejore destrezas y habilidades que le capaciten para abordar los contenidos de asignaturas de cursos posteriores de la titulación.

La masificación de los grupos de laboratorio de prácticas, con grupos de hasta 20 estudiantes por sesión son la mayor dificultad para la evaluación de esos trabajos prácticos. Esta dificultad es compartida por muchas otras asignaturas de la titulación que tienen una significativa componente práctica en sus contenidos, de modo que casi todas las cuestiones que en este trabajo queremos plantear son comunes a muchas asignaturas e incluso a otras titulaciones.

Una vez realizadas tres prácticas iniciales en las que

el alumnado se familiariza con el Sistema Operativo Linux, con el entorno de máquina virtual en el que desarrollará sus programas y con el editor `vim`, que es el que se utiliza inicialmente, todas las prácticas restantes abordan contenidos de programación que cubren los siguientes tópicos:

- Primeros programas y conceptos básicos
- Expresiones y tipos de datos
- Alternativas
- Iteraciones
- Funciones
- Cadenas de texto (`std::string`)
- `std::array` y `std::vector`
- Ficheros
- Introducción a la Programación Orientada a Objetos

Desde hace ya varios cursos en IB se viene usando la plataforma *Jutge.org* (juez) [8] para la evaluación de las prácticas. Se trata de una plataforma que ha sido desarrollada en la Universidad Politécnica de Catalunya, diseñada tanto para profesorado como para alumnado. La plataforma aloja una gran cantidad de problemas (aproximadamente unos 2100) que cubren diversidad de tópicos incluyendo entre ellos los fundamentos de la programación. Los problemas están perfectamente descritos y contienen un conjunto de tests que el código de usuario ha de verificar. Como una de sus características, *Jutge.org* hace hincapié en el trabajo del alumno, por lo que resulta especialmente útil para reforzar el enfoque de aprender haciendo, en nuestro caso, programando.

El modo de funcionamiento de *Jutge.org* requiere que cuando un estudiante resuelve un problema suba el código fuente de su solución a la plataforma ² donde se comprueba que sea correcta pasando los correspondientes tests, de los cuales algunos son públicos y otros privados. A una solución se le asigna el veredicto AC (*Accepted*) cuando pasa todos los tests existentes para el problema. En su cuenta de *Jutge.org* cada estudiante dispone de un cuadro de mandos en el que se recopila el número de envíos que ha realizado, el de problemas aceptados y rechazados así como diversos gráficos que muestran la evolución en su trabajo con los problemas de la plataforma.

Este modo de trabajo se aprovecha en las prácticas de IB: en cada sesión de evaluación se le pide al alumnado que resuelva un pequeño número de problemas (programas) de *Jutge.org*. La evaluación de la sesión depende no solo del número de problemas resuelto (no suelen ser más de dos o tres) sino de la calidad del código desarrollado. *Jutge.org* comprueba exclusivamente que el código evaluado funcione correctamente, mientras que en IB se enfatizan otros aspectos del código

²*Jutge.org* Home Page <https://jutge.org/>

que nos parecen tanto o más relevantes que el propio funcionamiento del mismo, que obviamente es un requisito ineludible. Muchos de los requisitos que se exigen a los programas de prácticas de IB se definen en la Guía de estilo de referencia que se sigue en la asignatura ³.

Relacionamos a continuación algunos de los requisitos exigidos, que obviamente se introducen al alumnado de forma progresiva:

- Correcto sangrado del código
- Adecuado uso de espacios y signos de puntuación en el código
- Adhesión a las reglas de nombrado de identificadores de variables, funciones, clases, etc.
- Que todos los identificadores utilizados (salvo excepciones puntuales) sean significativos, evitando el uso de “identificadores de un único carácter”.
- Los programas deben escribirse de forma modular incorporando diversas funciones en la solución.
- Que todos los ficheros, funciones y métodos del código incluyan un breve prólogo con comentarios en formato Doxygen exponiendo la información más relevante sobre el elemento (función, clase, fichero, ...) en cuestión.
- Que la compilación de todos los programas se automatice mediante el uso de herramientas como `make` o `cmake`.
- Que los parámetros de tipos estructurados que se pasen a una función/método estén sean referencias constantes.
- Que los métodos definidos en las clases de los programas sean *const friendly*.

Un problema recurrente en todas las asignaturas que requieren la evaluación de prácticas de programación es la infracción por parte de algunos estudiantes de las reglas de código de conducta que establecen que los trabajos presentados a evaluación han de ser programas originales realizados por sus autores. Ante la dificultad de acreditar fehacientemente esta restricción en una sesión de evaluación con un elevado número de estudiantes en el aula de prácticas y con un tiempo tan limitado se ha optado por valorar casi exclusivamente el trabajo que el estudiante realiza *in situ* en la sesión de evaluación, y no los ejercicios que durante la semana ha realizado en calidad de preparación para esa evaluación.

En junio de 2021, GitHub lanzó *Copilot* [7], un “programador de pares de IA” que genera código en diversos lenguajes a partir de cierto contexto como comentarios, nombres de funciones y código adyacente. *Copilot* se basa en un modelo que se entrena con código abierto [2].

³Google C++ Style Guide

<https://google.github.io/styleguide/cppguide.html>

En febrero de 2022 *DeepMind* publicó *Alphacode*, un sistema basado en Inteligencia Artificial (IA) que puede competir con un humano en la resolución de problemas sencillos de programación. Según resultados publicados en *Science* [6], *Alphacode* gana en un 50 % de ocasiones a humanos en competiciones de resolución de problemas de programación.

El pasado 30 de noviembre, *OpenAI* lanzó *ChatGPT* [9], un *chatbot* interactivo de propósito general basado en GPT-3.5 [5], un modelo lingüístico autorregresivo de tercera generación que utiliza *DeepLearning* para producir textos similares a los escritos por humanos.

Tanto *ChatGPT* como *AlphaCode* son “grandes modelos lingüísticos”, es decir, sistemas basados en redes neuronales que aprenden a realizar una tarea a partir de ingentes cantidades de texto generado por humanos. De hecho, ambos sistemas utilizan prácticamente la misma arquitectura, siendo la principal diferencia entre ellos el conjunto de datos con que son entrenados, lo cual los dirige a diferente tipo de tareas. Posiblemente debido a que se trata de un asistente de propósito general, capaz de discutir sobre materias tan dispares como Derecho, Filosofía, o Programación de ordenadores, ha sido el lanzamiento de *ChatGPT* el que ha suscitado un mayor interés no solo en la comunidad informática sino también en medios de comunicación de carácter generalista [3] atrayendo la atención de más de un millón de usuarios apenas cinco días después de su lanzamiento.

A pesar de que hay un consenso generalizado en la comunidad informática respecto a que, al menos en su estado actual, estos asistentes no van a suponer la desaparición de los programadores de ordenadores [1], también es incuestionable el papel que estos asistentes pueden llegar a tener en la toma de decisiones en diversos campos [4].

La siguiente sección aborda el impacto de *ChatGPT* en la realización de prácticas de IB.

3. Experiencias con *ChatGPT*

Todos los enunciados de problemas de *Jutge.org* están públicamente disponibles en la plataforma ⁴ y cada uno de ellos cuenta con una descripción clara y precisa así como un conjunto de tests públicos.

Aunque el estudio que se ha realizado no es exhaustivo, hemos utilizado *ChatGPT* para tratar de resolver diversos problemas de *Jutge.org*. Se ha intentado describir al bot los problemas tanto en inglés (idioma en que están descritos la mayoría de los problemas de *Jutge.org*) como en español. Para ello se ha iniciado un nuevo chat con el bot en uno u otro idioma dependiendo del caso. Si bien las soluciones que entrega el siste-

⁴Jutge Problems <https://jutge.org/problems/>

ma no son exactamente las mismas, no se han hallado diferencias significativas en cuanto a la calidad tanto algorítmica como de estilo en las soluciones obtenidas.

Para que el bot resuelva cualquiera de los problemas no ha resultado necesario suministrarle información adicional a la que figura en el propio enunciado de *Jutge.org*, habiéndosele entregado al bot la información correspondiente a la salida que el programa debiera entregar para los diferentes tests públicos de *Jutge.org*.

Por otra parte, si no se le especifica explícitamente, las soluciones que produce *ChatGPT* generalmente no siguen reglas de estilo o documentación específica, pero si en la especificación del problema se incluyen estas reglas como requisito, la solución cumple con los requisitos que se especifican. En general, para todos los problemas analizados se han indicado al bot instrucciones (en el idioma de la sesión) del tipo:

- Base su solución en una función que será invocada desde `main()`.
- El programa debe incluir comentarios de cabecera en formato Doxygen.
- Haga que el código cumpla con los requisitos de la Guía de Estilo de Google para C++.
- Haga que el programa utilice identificadores significativos para todas las variables.
- No utilice identificadores de un único carácter en el programa.
- No incluya en el programa comentarios innecesarios.

Es más, al tratarse de un bot interactivo, si la solución que entrega el sistema no es satisfactoria en algún sentido, se le puede pedir que refactorice el código para cumplir con algún requisito y en ese caso en general el sistema adecúa su respuesta para cumplir con el requerimiento.

Consideremos el problema *Fermat's last theorem* (1). La descripción que se entrega inicialmente a *ChatGPT*, copiada casi literalmente de *Jutge.org* es la siguiente:

A famous theorem of the mathematician Pierre de Fermat, proved after more than 300 years, states that, for any natural number $n \geq 3$, there is no natural solution (except for $x=0$ or $y=0$) to the equation $x^n + y^n = z^n$.

For $n = 2$, by contrast, there are infinite non-trivial solutions. For instance, $3^2 + 4^2 = 5^2$, $5^2 + 12^2 = 13^2$, $6^2 + 8^2 = 10^2$,

...

Write a C++ program that, given four natural numbers a , b , c , d with $a \leq b$ and $c \leq d$, prints a natural solution to the equation $x^2 + y^2 = z^2$ such that $a \leq x \leq b$ and $c \leq y \leq d$. The program input consists of four natural numbers a , b , c , d such that $a \leq b$ and $c \leq d$.

The program output should be a line with a natural solution to the equation $x^2 + y^2 = z^2$ that fulfills $a \leq x \leq b$ and $c \leq y \leq d$.

If there is more than one solution, print the one with the smallest x .

If there is a tie in x , print the solution with the smallest y .

If there are no solutions, print "No solution!".

For example, if the input is: 2 5 4 13

The output should be

$3^2 + 4^2 = 5^2$

and if the input is: 1 1 1 1

The output should be No solution!

y con ella, el bot entrega una solución que es aceptada (AC) por *Jutge.org* pero que no cumple algunos de los requisitos que se exigen en IB. Después de sucesivas interacciones con el bot en el que se le indica:

1. Could you please avoid the use of single character identifiers and use meaningful names instead?
2. Please, use a function called from `main()` in your solution
3. Can you make your code compliant with the Google Style Guide for C++?
4. Can you avoid the use of "using namespace std"?
5. Can you write a space on both sides of any binary operator, as the Google Style Guide for C++ requires?
6. Can you finally include Doxygen format header comments in the code?

Se obtiene un código que es aceptado por *Jutge.org* y que cumple los requisitos exigidos al alumnado de IB en la evaluación de sus prácticas.

Así por ejemplo, en el problema *Fermat's last theorem* (2) se especifica "Input has several lines, each one with four natural numbers a , b , c , d such that $a \leq b$ and $c \leq d$."

Consideremos a modo de ejemplo el problema *Primality*. (P48713). Se trata de determinar si cada uno de los números naturales de una secuencia es o no primo. Si el enunciado del problema, junto con los tests se le pasan a *ChatGPT* en inglés, tal como figuran en *Jutge.org*, la solución que obtiene es la más habitual para un programador inexperto, la de fuerza bruta consistente en probar todos los divisores en el rango $[2, N-1]$. Esta solución no recibe el veredicto "Accep-

ted” en *Jutge.org* sino que la plataforma indica como veredicto *Execution Error (time limit exceeded)*. Ello se debe a que espera un algoritmo más eficiente para este cómputo.

Esta solución es la que cabría esperar de un estudiante de primer curso de informática. Con frecuencia observamos estudiantes que ofrecen un algoritmo óptimo y ello es una pista para detectar que, posiblemente han hallado la solución en algún foro. Al preguntar al estudiante la razón por la que no recorre todo el rango de búsqueda cabría esperar una respuesta en la que el estudiante indique que ha investigado el problema y aprendido sobre el mismo, pero es frecuente una respuesta del tipo “*lo he probado y he observado que funcionaba*”, que obviamente no se considera adecuada para la evaluación de un ejercicio práctico.

En el propio enunciado del problema se indica una pista para un algoritmo de menor complejidad. Si a *ChatGPT* (en este caso en inglés) se le indica *Could you optimize the is_prime() function for a better performance?* el bot modifica la función, entregando en este caso una versión que sí es aceptada en *Jutge.org*. A pesar de haberlo indicado en el enunciado del problema, la función entregada no cumple con el convenio que establece la Guía de Estilo de Google, pero si se le indica esa circunstancia, el bot corrige el identificador de la función. El código completo de la función es el que se muestra en el Listado 1.

En el caso de *Primality*, en una sesión diferente con *ChatGPT* se le ha entregado el enunciado en español y en ese caso, el bot entrega directamente la solución con la optimización del código, pero el código no cumple con el estándar de Estilo en cuanto a la colocación de las llaves de apertura y cierre de bloques en C++.

4. Conclusiones

Si bien la llegada de los asistentes basados en IA a las aulas no implicará que el profesorado deje de ser necesario, sí es cierto que estas tecnologías van a impactar en la praxis docente y hemos de adaptar nuestras metodologías para incorporar estos cambios.

Algunas preguntas que toda lo expuesto anteriormente motiva y sobre las cuales debemos reflexionar son las siguientes:

- ¿Debemos ignorar la existencia de los asistentes

basados en IA o debemos por el contrario, incorporarlos a nuestra práctica docente?

- ¿Hemos de modificar nuestras metodologías docentes?. En caso afirmativo, ¿cómo y qué cambios debiéramos introducir?

Referencias

- [1] Davide Castelvetti. Are ChatGPT and AlphaCode going to replace programmers? *Nature (London)*, 12 2022.
- [2] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, y Greg Brockman. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021.
- [3] Jordi Pérez Colomé. “funciona muy bien, pero no es magia”: así es ChatGPT, la nueva inteligencia artificial que supera límites. *El País*, diciembre 2022. <https://elpais.com/>.
- [4] Tiffany H. Kung et al. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. En *NPJ Digit. Med.*, 2022.
- [5] Luciano Floridi y Massimo Chiriatti. Gpt-3: Its nature, scope, limits, and consequences. *Minds Mach.*, 30(4):681–694, dec 2020.
- [6] Yujia Li, David Choi, Junyoung Chung, Nate Kushman, y Julian Schrittwieser et al. Competition-level code generation with AlphaCode. *Science*, 378(6624):1092–1097, 2022.
- [7] Nhan Ton Nguyen y Sarah Nadi. An Empirical Evaluation of GitHub Copilot’s Code Suggestions. *2022 IEEE/ACM 19th International Conference on Mining Software Repositories (MSR)*, pp. 1–5, 2022.
- [8] Jordi Petit, Josep Carmona, Jordi Cortadella, Jordi Duch, Giménez Omer, Anaga Mani, Jan Mas, y Enric Rodríguez-Carbonell. Jutge.org: Characteristics and experiences. *IEEE Transactions on Learning Technologies*, 11(3):321–333, 2018.
- [9] Xingxing Zhang, Jinchao Zhang, y Maxine Eskenazi. ChatGPT: A task-oriented dialogue system based on pre-trained language model. *arXiv preprint arXiv:2012.16641*, 2020.

Listado 1: Test de primalidad suministrado por *ChatGPT*. Complejidad $O(\sqrt{n})$

```
/**
 * @brief Function that determines if a given number is prime or not.
 *
 * @param number The number to check for primality.
 * @return true if the number is prime, false otherwise.
 */
bool IsPrime(int number) {
    if (number <= 1) return false; // 1 and 0 are not considered prime.
    if (number <= 3) return true; // 2 and 3 are prime.
    if (number % 2 == 0 || number % 3 == 0) return false; // Check for divisibility by 2 and 3.
    for (int i = 5; i * i <= number; i += 6) { // Check for divisibility by numbers of the form 6k +/- 1.
        if (number % i == 0 || number % (i + 2) == 0) return false;
    }
    return true;
}
```